Proceedings of the Sawtooth Software Conference

"Gaining A Competitive Advantage Through PC-Based Interviewing and Analysis"

VOLUME I

1989

FOREWORD

We're pleased to make available these written versions of the presentations given at our third Sawtooth Sawtooth Conference, in Sun Valley, Idaho (June 1989). The speakers addressed issues relating to computer interviewing, conjoint analysis, perceptual mapping, and cluster analysis - with a look toward using these techniques to "gain a competitive advantage."

These papers add substantially to the practical and theoretical literature, and we thank the authors for their contributions. The papers are in the words of the authors; only light copy editing was performed.

We want to maintain an open forum at our conferences; we welcome readers' comments about topics that would be of interest to them at our future conferences.

Margo Metegrano
Editor, Volume I
September 9, 1989

"Gaining A Competitive Advantage Through PC-Based Interviewing and Analysis"

THE DISK-BY-MAIL SURVEY AS A COMPETITIVE TOOL

PRACTICAL ISSUES IN COMPUTER INTERVIEWING

CONJOINT ANALYSIS AS A TOOL FOR COMPETITIVE PRICING

EMERGING TECHNOLOGY

COMPUTER INTERVIEWING

PERCEPTUAL MAPPING

CONJOINT ANALYSIS

CONJOINT ANALYSIS (Continued)

CLUSTER ANALYSIS

Disk-By-Mail Surveys: Three Years' Experience

Brant Wilson
COMPAQ Computer Corporation

At COMPAQ Computer Corporation, we have been using disk-by-mail (DBM) surveys now for over three years. There is a very simple reason for this: We believe in them!

Our goal is to stay ahead of our competitors and so we are apt to consider many new research approaches that have a reasonable chance of success. When we learned of this new technique in 1985, we were very interested. Of course, we had a few concerns as we would about any new approach. However, once we were able to try disks-by-mail we were hooked - because it really worked! In fact, it worked even better than we had expected.

In looking back on our willingness to "test" this new approach, three major factors were central. First, it represented advanced technology and an ability to be leading edge in our orientation. This is very important because it is consistent with COMPAQ's image and the way we do business. Second, we thought our customers would like it. Here again, if it is good for our customers, chances are it is good for us. Finally, it appeared to offer some real benefits in our research effectiveness. To help us "work smarter" so to speak, and perhaps, to even improve on data quality. If that was not enough to get us started, then nothing would be.

How We Have Used DBM Surveys

The first study we tried on disk was an annual brand image survey in March 1986, which we had first conducted in 1985. It was well-suited for the technique because one key segment, MIS Managers, was almost certain to have access to an IBM-compatible PC. Also, we were intrigued to see if the diskette would increase the response rate for this over-studied group.

To test this premise, we designed an experiment whereby each respondent was sent both a diskette and a pencil-and-paper questionnaire. An enclosed letter explained that either could be returned. Among this group, over twice as many returned the diskette! By this we knew we were on to something special, and we have been using the diskette alone for this survey ever since.

Another type of application for disks-by-mail is research on how potential product benefits or capabilities fit the user's application. In one particular survey, for example, we had general knowledge of users' applications and preferences from focus groups. We then wanted to quantify and extend this knowledge - but many of the concepts or capabilities we wanted users to evaluate were much too complex for a standard survey.

One example of such a capability is multi-tasking, the idea of a computer conducting more than one operation at a time. We asked respondents about their desire for this capability by first asking a number of questions describing their current PC usage, then asked about what other things they would like their PC to be able to do. Finally, allowing the diskette survey to base a final set of questions on stored information from the previous questions, we asked them to rate their preference for a narrower set of capabilities, which included multi-tasking. As a result of the special ability of the diskette based approach to build intelligent respondent profiles and complex skip patterns, we were able to ask specific questions to only those users for whom the questions were relevant. In this manner we got information which otherwise would only have been available through a more expensive phone survey using trained interviewers.

Finally, a very important use of the DBM is for our owner registration. This is a high priority area for COMPAQ, and we were already achieving what we considered to be above average response rates with a card version. The question was, can we improve the response rate and the quality of information with a diskette based approach? And, would it be affordable? Well, it has proven better in both areas. Our response rate is twice as high for the disk and we now receive several times more information from each customer. The cost of including a diskette in every product is indeed more expensive, but it has proven worthwhile and we are narrowing the cost gap through component cost reductions and economies of scale made possible by high product volume.

Some of Our Concerns About DBM

Initially, we felt that even if successful in response rate, DBM surveys would have a sample bias due to having greater appeal to "sophisticated" user segments such as engineers. There is not a clear cut answer to this problem. We do know that these groups are already more likely to respond to any survey, and our surveys have not shown that they make up a significantly larger percent of diskette-based respondents than they do on other studies. So we are less concerned today, but this issues bears watching, especially if the respondent group is more "mainstream" than our target customers.

As with most technologies, standards emerge and they evolve. Thus, we wondered if (gradual) shifts to newer disk media (e.g. 3 1/2"), operating systems (e.g. OS/2), and/or user interfaces (e.g. Windows) would cause early commitment to DBM surveys to be an unwise investment. These changes will be larger issues over time, but certainly do not destroy the value of our investment in the approach.

Costs of DBM surveys are generally higher than for pencil-and- paper surveys so justifying this cost is a challenge. The up- front development work (programming) and cost of mailing is higher, but is partially balanced by lower back-end costs of coding and data entry. Certain costs which were exorbitant but are now dropping (e.g. diskette mailer envelopes) are starting to reduce this gap, and large volume mailing helps reduce it even more. Ultimately, management must be convinced that the research obtained from a disk-by-mail survey is worth paying a little bit more for in the short term due to the greater long term benefits of the higher-quality research obtained.

In any field, but especially in technology markets, management needs answers today, not tomorrow. Thus, any mail survey causes initial skepticism because by the time results are in, they may not be current with market conditions, or on deadline. Diskette- based surveys do not solve this problem, but they have tended to draw an earlier pattern of returns, with a shorter data entry process, allowing some time savings. Beyond that, we work within normal mail survey constraints and should not promise a dramatic difference in total study time.

We noticed that doing DBM surveys places a degree of pressure on staff expertise in two particular ways. However, there may be others we have not noticed. First, even if Ci2 (Ci2 System for Computer-Interactive Interviewing) programming is contracted outside, someone on your staff should gain some expertise in it in order to carefully manage this effort. This expertise is not extremely difficult for someone to acquire, but should be considered in staff time allocation. Second, because of the tremendous opportunities which DBM offers for getting more specific and useful information, those writing the questions must possess the ability to take advantage of these benefits. In other words, ability to get higher quality information from the approach is only available to the skilled question writer. While true for all surveys, this situation is accentuated for DBM surveys. Finally, one concern in justifying DBM surveys to management and to ourselves was the noted linkage of the approach to conjoint techniques. That is, unless a conjoint study was being planned, would the DBM approach offer anything new? Well, the above comments say that it does, but this needs to be reinforced when selling the benefits to those less familiar with these benefits and more likely to simply associate DBM with conjoint.

<u>Benefits and Opportunities</u>

When one looks at the continued growth in the installed base of industry standard personal computers, the opportunities for surveying this audience is astounding. As current estimates show, the audience for DBM surveys is already large and is expected to grow consistently at least through the 90's. Beyond that, penetration of PC's in the general business populace is also growing and represents an increasingly high-incidence user audience even among this broader group.

As stated above, the registration of owners or customers clearly is a major opportunity to use the DBM approach. As overall business penetration increases, this makes the approach more feasible for non-computer categories.

Response rate increases like the ones we have seen are a commonly touted benefit of the DBM approach, as has been well documented. However, I believe major opportunities could still exist for even greater increases in response rates by creative use of incentives in conjunction with various diskette-based survey approaches. On the downside, we will need to be clever to sustain consistent response rates after some of the "novelty" of this approach wears off. (By the way, one could argue that the research community is better off if recent diskette-based approaches to other marketing and sales activities do not take off.)

The DBM survey provides an opportunity to ask complex questions simply. What is required is some creative thinking in developing the questionnaire--learning what the technology allows and then exploiting it. The benefits are enormous in getting information that was previously unavailable from mail surveys. This is a major benefit making DBM unique among mail survey approaches or other self-administered techniques. Quality of information is a significant benefit of DBM surveys for several reasons. First, as just noted, it allows getting at information previously unavailable to us. Second, with higher response rates, we feel more confident that the population is being represented - thus our measures are more accurate from a statistical standpoint. Finally, we know respondents are enthusiastic about filling out DBM surveys. Happy respondents are more diligent, more accurate, and less likely to give fraudulent answers - all of which improve data quality.

For COMPAQ, an additional benefit of DBM surveys is that it represents a leading edge approach to information gathering and thus appeals to our image of ourselves--a company which uses advanced technology as well as it manufacturers and markets it. We want our customers to feel that we are offering them the latest, even when we send a them survey.

Summary

In summary, diskette-by-mail surveys are becoming more common, and for good reason. Though sharing many problem areas with other survey techniques, they offer many benefits and opportunities, both today and in the future. But most of all, they really work! And if properly used, they can offer a real competitive advantage.

Customer Satisfaction Research Using Disks-By-Mail

Peter Zandan and *Lucy Frost*
IntelliQuest

Introduction

Over the past three years, IntelliQuest has conducted 34 disks-by-mail (DBM) surveys and mailed over 30,000 diskettes. These projects were administered nationally and internationally, with telephone pre-recruitment and without, to respondents in different job functions and with varying levels of computer expertise. Through this experience, we have concluded that in the appropriate situation, disks-by-mail can be the preferred data collection method. (Pilon and Craig, Sawtooth Software Conference Proceedings, 1988).

This paper is designed to assist those who are considering using the disks-by-mail method of data collection. It addresses key issues in conducting disks-by-mail surveys and provides an illustrative case study.

I) Disks-by-Mail Success Factors

Success factors for disks-by-mail surveys are similar to those for conventional paper mailed surveys. Appropriate pretesting, reasonable questionnaire length, clearly phrased questions, mailing list quality, effective cover letters, respondent incentives, survey packet appearances, first class stamps, second mailings of the survey, pleasant questionnaire format and design, use of personalized salutation, etc. all contribute to the success of the research regardless of survey modality. The following four factors, however, are unique to DBM surveys.

1) Access to Personal Computers

The key factor for determining the feasibility for a DBM survey is the targeted sample's access to personal computers. With some groups this is a simple prerequisite. For example, computer hardware and software companies that want to research their own and/or competitors' customers can be confident that members of the target group are computer users.

For those survey groups who do not necessarily have easy assess to a computer, careful consideration and pre-screening by the researcher is required. Research firms that follow personal computer adoption (Dataquest, Gartner Group) estimate that 75% of the people who work in offices have access to PC-compatible computers. For example, most purchasing agents or financial analysts in Fortune 1000 companies have access to personal computers. However, these numbers vary depending on the type of job function, industry type, and company size. For example, 98% of those firms with 500 or more employees have one or more personal computers, while only 25% of the firms with fewer than ten employees have at least one personal computer. As the proliferation of personal computers continues, (it is estimated that there were over 10 million new personal computers shipped in the U.S. in 1988), the opportunity for conducting DBM surveys increases.

Residential markets present a much more difficult group for using disks-by-mail surveys. Because of the lack of standards and the low penetration rate (1 out of 5 households has a computer), DBM is generally not an appropriate data collection method for residential markets.

2) Diskette Duplication and Quality Control

All disk drives are not alike. When preparing diskettes for duplication, be sure to have a high-quality aligned disk drive. If your master diskette is on a system that is even slightly out of alignment, this can cause problems with a respondent's ability to use the duplicated survey disk. The solution to this is to use professional disk duplication rather than using an available computer in your office. We also recommend randomized testing of 5% of the duplicated diskettes to ensure that the diskettes are working properly. Even when these methods are employed, expect that a small percentage (1 %) of the diskettes will be damaged.

3) Respondents' Computer Systems

Four system-compatibility areas need to be kept in mind in order to conduct a successful survey.

a) Operating systems

Although IBM compatibility is an acknowledged standard for personal computers, over 10% of the personal computers in use are Macintoshes, which are incompatible with the IBM standard. Pre-screening respondents to identify those who use Macintoshes can assist in compensating for this inconsistency. For example, we have sent IBM-compatible diskettes to Macintosh users, apologizing for the inconvenience and providing a substantial incentive for completing a survey on another machine. By specifically encouraging their participation, the response rate for Macintosh users was the same as for IBM-compatible users.

b) Monitors and color

Using colors on the survey screen can create reading and contrast problems for those respondents with monochrome monitors. Some color monitors also display certain colors in distinctive shades, making questions unreadable due to contrast. Besides pretesting the programmed diskettes on a number of monitor types, we recommend using high contrast colors when programming with color.

c) Disk drives

Incompatible disk drive sizes and types can cause problems and inconveniences for respondents. For example, the two standards of 3 1/2 inch and 5 1/4 inch make it imperative to send both types of diskettes to a respondent unless there is pre-screening to determine the size of disk drive they use.

Another problem is that diskettes containing files created by high density drives on AT or 286 class machines cannot be read by most PC or 8086/8088 class machines (which represent over half of the existing PCs in use today). Thus, high-density drives should not be used to write on diskettes for distribution.

d) DOS operating systems

Current copyright laws require that you do not mail DOS on your diskettes. Make sure you do not format disks with the /s ("system") option or copy the command.com file onto diskettes to be mailed.

4) Technical and Survey support

Our experience has shown the importance of providing survey and technical telephone support for respondents. The value of survey phone support is to assist people who have system incompatibilities, who have received defective disks or disks that are damaged in the mail, and most importantly, people who need help with the technical aspects of taking the survey.

Fewer than 5% of the recipients of DBM surveys use our "800" telephone line for technical support. It is primarily used by novice computer users who appreciate the extra assistance.

II) CASE STUDY: Customer Satisfaction Research Using DBM

A corporation providing products for computer users wanted assistance in developing a diskette-based client satisfaction survey. The corporation had been conducting a semi-annual satisfaction survey of 300 customers and ultimately wanted to expand it to include 2000 new customers each month. Because of costs and questionnaire length, the corporation's research department opted to replace the phone data collection approach with a diskette mailing to all of its customers. The anticipated benefits, in addition to monitoring customer satisfaction, included:

1) the opportunity to capture customer information to update their corporate-wide customer database (that would assist them in their marketing and sales decision making.)

2) the opportunity to enhance their image as a technological leader through the use of this innovative, technologically- advanced survey mode.

The results reported here are from the first wave or pre-test of the survey. To ensure the quality of the project, the client wanted to assess their customers' attitudes regarding the survey technique and also examine non-response bias. The following describes the six phases of the research project and the findings.

1) Questionnaire Design

The survey instrument was designed to capture customer satisfaction, purchase process issues, product usage and customer demographics. A total of 1,311 disks were mailed to 951 different sites, one disk for each product recently purchased. The disk was accompanied by a cover letter signed by the company president on corporate letterhead.

The survey instrument was reviewed and approved by market research, sales, marketing and direct marketing, and MIS representatives.

The survey contained approximately 185 variables, and respondents took between 25 and 35 minutes to complete it.

2) Fielding Issues

The questionnaire package was mailed first class. On the basis of information provided by the survey sponsor, we enclosed a diskette that was compatible with the potential respondent's personal computer. Customers were not notified in advance of receiving the survey either by mail or by telephone screening. We offered an "800" telephone number for individuals to call if they had questions. We received approximately 31 calls. Exhibit A shows the nature of the calls:

Exhibit A

800 Number Response

- **31 Total Calls**

  15 -- Disk usage information

  9 -- Disk failure

  7 -- Disk destroyed in mail

3) Response Rate

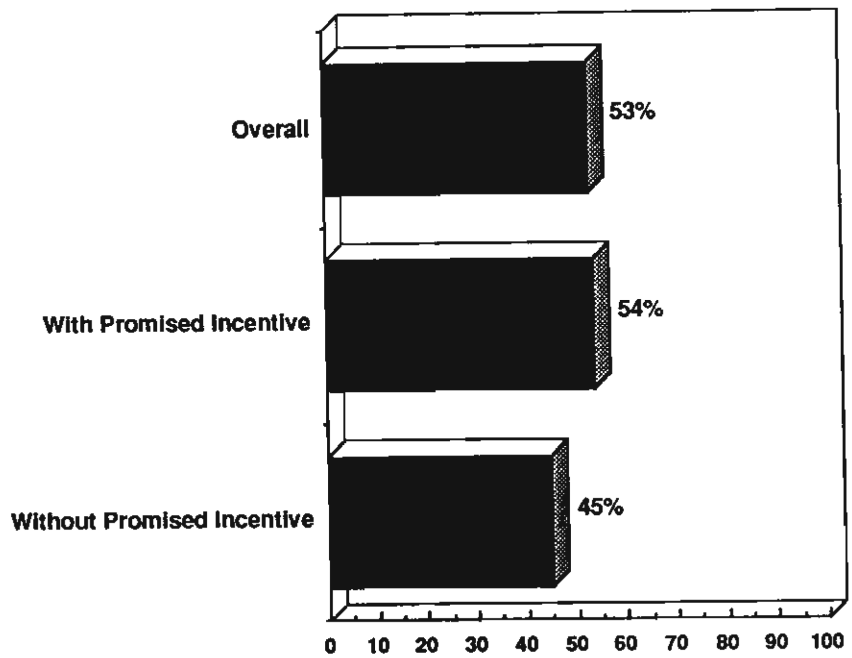
A total of 951 sites were sent diskettes. Diskettes were returned over a period of five and a half weeks. Completed disks were returned from a total of 500 sites, producing a response rate of 53%. (See Exhibit B)

4) Incentive

To examine the effect of offering an incentive, potential respondents were randomly assigned to one of two groups. The incentive group, 85% of the sample, was promised an incentive (a coffee mug) for responding. The no-incentive group, 15% of the sample, was offered no incentive. The promised incentive increased the response rate by approximately 9%, from 45% to 54%.

Exhibit B

Response Rates Based on Cooperative Sites Who Complete & Return At Least One Disk



5) Debriefing Survey

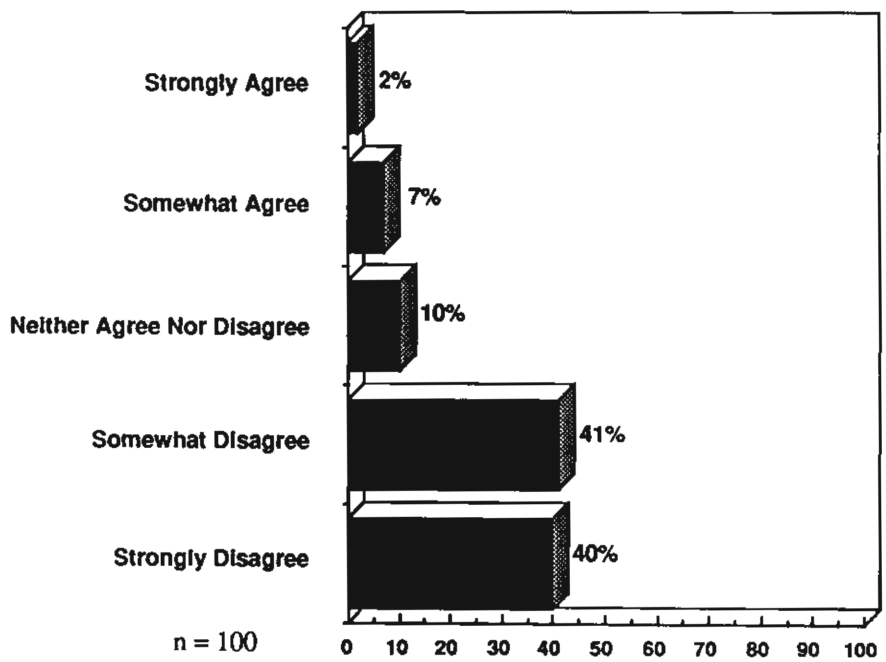
One hundred of the respondents who completed and returned the diskette-based survey were randomly selected to be interviewed over the telephone to assess their opinion of the survey.

The key findings from this survey are:

1) The vast majority did not find that the survey was an "unwelcome nuisance."

Exhibit C

How much do you agree or disagree with the statement:
"It is an unwelcome nuisance to get this survey."

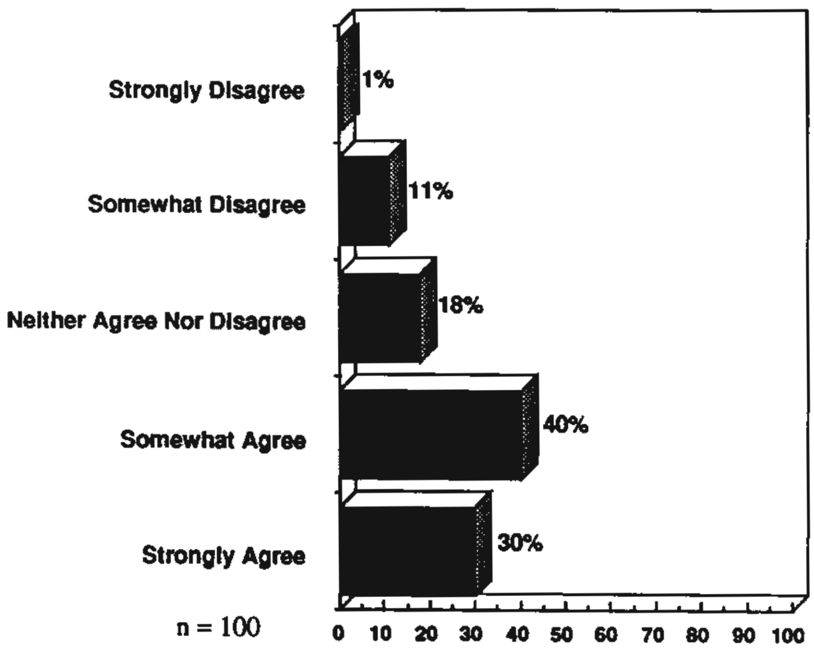


2) Most respondents were generally impressed that the survey sponsor was conducting the survey.

Exhibit D

How much do you agree or disagree with the statement:

"I'm extremely impressed; no other company cares this much about their customers."

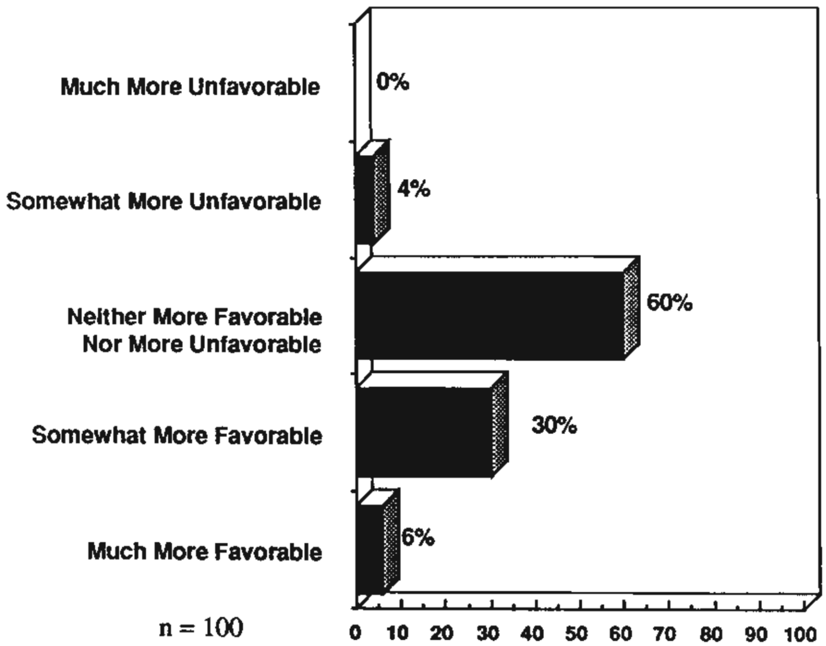


n = 100

3) Over 35% of the respondents actually stated that their attitudes improved toward the sponsoring company since receiving the survey.

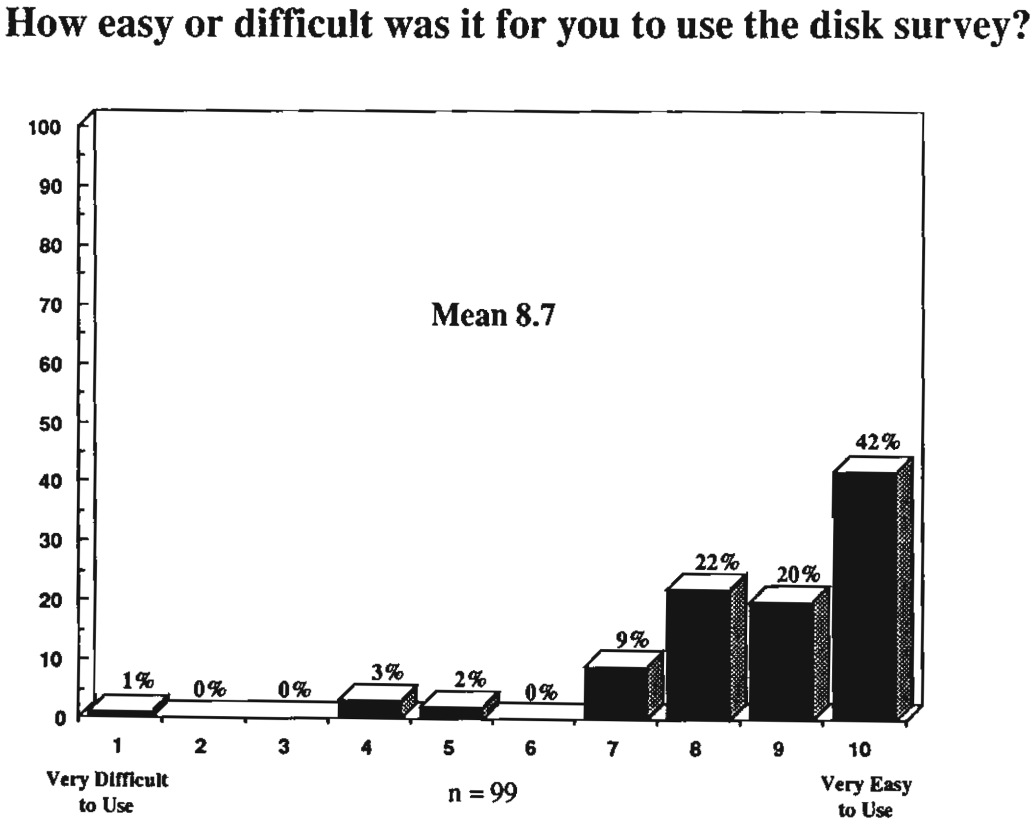
Exhibit E

Since receiving the survey, have your attitudes towards XYZ in general become...



The vast majority found the DBM survey easy to use.

Exhibit F



Respondents perceived that the survey took them less time than was actually measured electronically by the survey disk. On average, they reported that the survey took 19 minutes. On average, the clock on the survey disk registered 33 minutes (outliers were eliminated in computing the average).

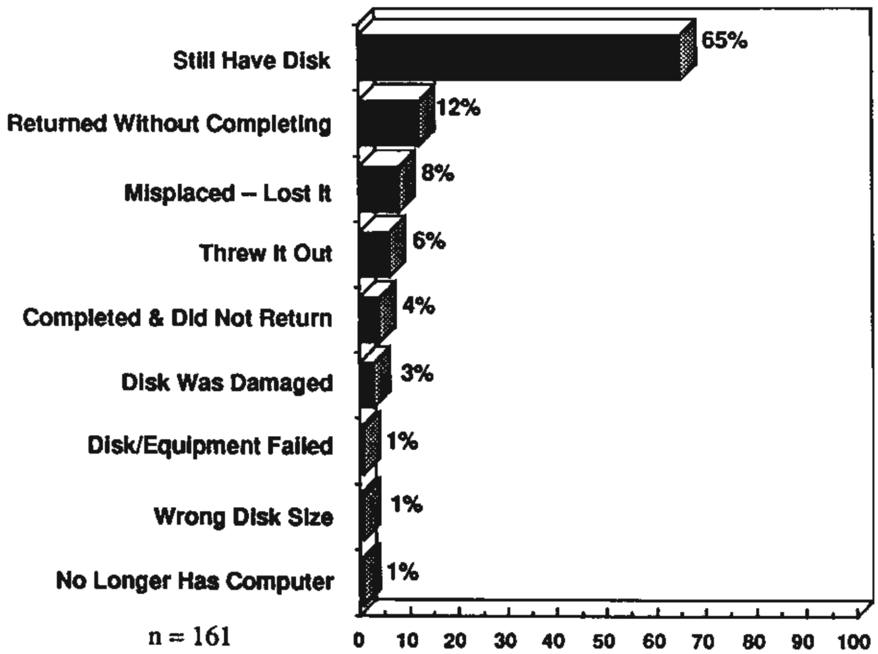
6) Non-response Bias

One hundred and sixty-five individuals who did not respond to the initial customer satisfaction survey were contacted by telephone to determine why they did not complete and return the diskette. In addition, they were asked demographic and satisfaction questions obtained in the original survey. This made it possible to compare and contrast respondents and non-respondents.

When asked "Can you tell me what you did with the disk based survey you received?," the majority of non-respondents stated they still had the disk.

Exhibit G

Could you tell me what you did with the disk-based survey you received?

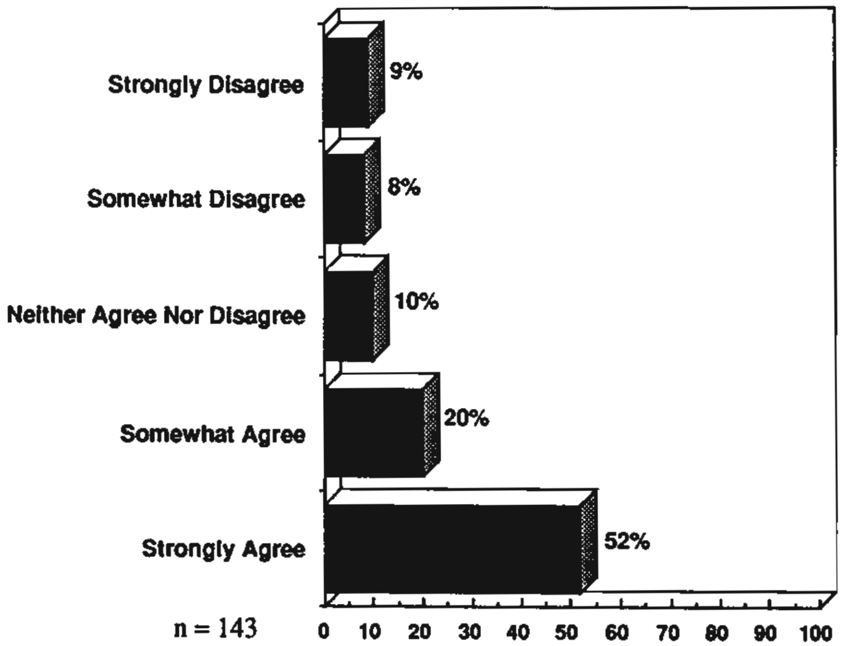


Almost 75% of the respondents said they just did not find time to complete and return the survey.

Exhibit H

How much do you agree or disagree that the statement may have contributed to preventing the disk from being completed and returned:

"I just did not find time to complete and return the survey."



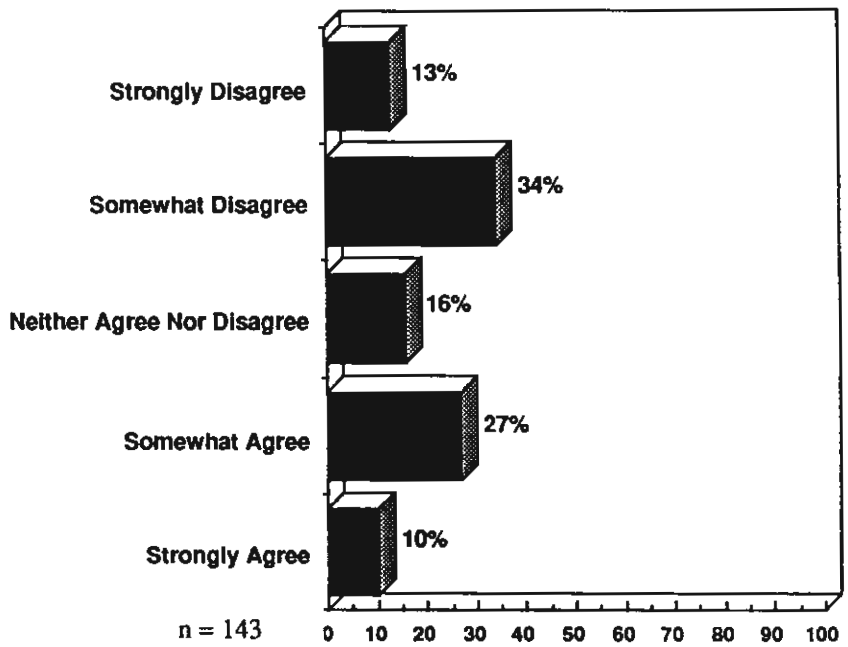
n = 143

Ten percent of the non-respondents said they had some fear that the diskette might have carried a computer virus. Thirty-seven percent of the respondents said that they generally do not participate in surveys.

Exhibit I

How much do you agree or disagree that the statement may have contributed to preventing the disk from being completed and returned:

"I generally do not participate in surveys."



Comparisons of respondents and non-respondents revealed only two differences.

1) Non-respondents stated that they were more likely to think the sponsoring company was trying to "sell something" through the survey than those who did respond.

2) Respondents were more likely than non-respondents to express the belief that the survey would have a positive effect on the sponsoring company's customer relations.

In general, very few differences exist. For example, differences between respondents and non-respondents were not found with regard to the following areas: overall satisfaction with specific areas of the sponsoring company, job titles, gender, income, education, age, company sales, and number of employees in organization.

III) Conclusion:

The major area of concern for the research community in evaluating DBM is its impact on the reliability and validity of the data collected. Does this method do a better job of measuring what it is intended to measure? As of today, this question has not been answered unequivocally. We need more experiments comparing conventional data collection methods with this new method. Only with such experiments can we determine whether computer-aided interviewing is really superior to the more familiar research methodologies.

Finally, no one should expect that conventional methods of data collection will be "replaced" by DBM. Successive information technologies have not replaced previous media. Printing has not replaced handwriting. Teleconferencing has not replaced face-to-face meeting. Television has not replaced radio. New technologies bring some new advantages but seldom eliminate the "old ways" of doing things. However, the new method tends to change the function of the old technology. Conventional techniques will certainly continue to be used in appropriate situations. Nevertheless, as more people use computers, DBM will almost certainly become a standard tool in market research.

CONTEXT-SPECIFIC CHOICE EXPERIMENTS FOR MULTI-FEATURED PRODUCTS: A DISK-BY-MAIL SURVEY APPLICATION

Shari Gershenfeld and *Terry Atherton*, Cambridge Sytematics, Inc.;
Moshe Ben-Akiva, MIT; and *Larry Musetti*, AT&T Business Markets

Introduction

Our presentation concerns the application of a disk-by-mail survey to conduct context-specific choice experiments for multi- featured products or services. The specific research application which will be described was done for AT&T by Cambridge Systematics. The software used for the study was developed by the Hague Consulting Group in Holland and is available through Cambridge Systematics in the U.S.

We begin by describing the key features of our approach for measuring preferences for complex, multi-featured products. We then present the AT&T research as a case study that illustrates how and why this approach is particularly well-suited to a disk-by-mail survey.

Methodological Approach

In contrast to standard "off-the-shelf" products (packaged goods for example), many products and services are sold customized to the needs or specifications of a particular person, household or firm. Many high technology products fit this description. They tend to have many optional features and thus to have a large number of potential attributes. One example would be a PBX telephone system which is sold with the size and feature specifications that meet a particular company's needs and which is priced according to those specifications. Another example is an automobile which is priced and sold with a different set of options or features for each person. A personal computer may also have different features, e.g. the amount of memory, disk space, etc. depending on the particular needs of the buyer. The new array of optional telephone calling features transform ordinary household telephone service into a multi-featured product. Many other examples of these types of products and services could be cited as well.

For products and services of this nature, pricing tends to be complicated, since the total price of the product depends on the particular features and product specifications required by each buyer.

The commonly used methods of collecting conjoint data are not particularly well-suited for such complex products. The challenges in collecting and modeling customer preferences for these products concern the relevance of the alternatives presented to respondents and the ability to realistically represent complex pricing structures in the survey design. Effective research requires a flexible approach to survey and experimental design and flexible survey software to incorporate those designs.

Our approach to measuring preferences for these products is to conduct context-specific choice experiments. The key feature of these choice experiments is that they mimic reality as closely as possible. The alternatives presented to a respondent are always full-profile in order to keep the choice task realistic. The respondent is asked to choose between two or more alternative products or services which are described with their full range of attributes; a respondent is never asked to express preferences between bundles of attributes or to indicate which attributes he or she considers important.

The alternatives which are presented are "context-specific" in that they are tailored to the relevant needs and characteristics of each respondent. The alternatives are not adapted on the basis of subjective criteria expressed by the respondent. In other words, the respondent is not explicitly asked to eliminate any attributes or attribute levels.

Consider an example of purchase preferences for automobiles. The respondent is not asked which automobile makes and sizes, for example, he or she would consider or not consider buying. Instead, in our approach, the respondent is asked about the car's intended uses and about the household's characteristics. This information is then used to construct realistic, relevant alternatives. If one of the car's intended uses is to transport the family and the family has seven members, for example, the respondent would not be presented with a subcompact car. This approach, i.e. placing a realistic alternative in a relevant context, leads to a more accurate and valid expression of respondent preferences.

Experience with commonly used conjoint methods has shown that respondents often resort to using short-cut decision protocols to complete survey tasks. One example of this phenomenon is known as the "prominence hypothesis" in which the respondent focuses on a particular attribute as the basis for his or her expression of preferences among alternatives. The artificiality of these experiments tends to detract from the relevance of the responses to actual market situations.

In our approach, we place the choice task in a realistic framework in order to make it familiar and meaningful to respondents. In this situation, it is more likely that the decision protocol that would be used in an actual choice situation will also be used in the survey choice situation.

These elements of our approach are applicable to any products for which preference data need to be collected. We are particularly interested, however, in products with optional features that are customized to a buyer's needs. These products tend to have complex pricing structures and a large number of attributes which are not necessarily ordinal nor independent of each other.

Again in order to maintain realism, the classical experimental design approach with emphasis on orthogonal designs is not feasible. Instead, we begin with a large fractional factorial design using all the attribute possibilities with the exception of price. The purpose of this design is to ensure sufficient variability in the data. We then apply a sequence of respondent-specific restrictions to eliminate irrelevant alternatives. The next step is customize the remaining alternatives according to respondent-specific requirements. Finally, we apply pricing algorithms that account for the specific features and product configuration included in the alternatives. We also account for the actual inter-relationships between features or attributes and thus avoid insensible combinations of attributes. Using the automobile example again, it is not realistic to ask someone to express a preference about a $10,0000 American-made sedan without specifying in more detail the make, model, engine size and other optional features. In reality, cars are purchased with these options and features and they are correlated with the price of the car. We allow for these correlations in our experiment.

AT&T Case Study

We will describe an application of this methodological approach in a disk-by-mail survey framework. The application concerned a market research study of network management systems (NMS). Network management systems are "high tech" multi-featured products.

In preparation for the introduction of a family of network management system products, AT&T needed to:

- Assess the total market for network management systems and AT&T's share of this market over the next 5 years
- Determine customer preferences among alternative packages of network management products, features and applications
- Measure willingness to pay for proposed products, features and functionalities
- Characterize customer market segments

Cambridge Systematics was hired to conduct research that would meet these objectives. Our approach involved the use of a disk-by-mail survey to collect preference information from telecommunications and data communications managers at large corporations. These data were used to develop discrete choice models that predict the decisions of corporations regarding the purchase of network management products.

The decision to use disk-by-mail surveys was influenced by the commonly cited advantages of this type of survey. For example, the targeted population - telecommunications and data communications managers at large corporations - is very heavily surveyed. Therefore, an innovative survey approach was needed to capture and maintain respondent interest.

The complexity of the product required that the survey be made extremely visual for the respondent in order to ensure comprehension; the complexity of the experimental and survey design suggested that a computerized questionnaire should be used to reduce the potential for interviewer error. However, the respondents were geographically dispersed, making in-person interviews (with laptop computers) prohibitively expensive and adding to the appeal of disk-by-mail surveys.

In addition to the "routine" advantages of the disk-by-mail surveys, it was an ideal survey modality through which to design and conduct complex interactive choice experiments.

In accordance with the approach described above, the survey began by asking for an extensive inventory of equipment and other information pertaining to the corporation's existing voice and data communications networks. Examples of survey screens in which this type of information was collected are shown in Figures 1 and 2.

This information was used to place context-specific restrictions on the network management system alternatives that would be presented in the following section of the survey. For this study, those restrictions related to the network elements which would need to be managed by a network management system and the manufacturers of the existing network elements for which connectivity would have to be provided. Open-ended responses could also be entered and included as attributes in the experimental design. For example, if the manufacturer of the corporation's PBX was not included in the list of PBX manufacturers, the respondent could type in the make of the firm's PBX; this manufacturer name would then appear later as one of the attributes of the network management system alternatives.

The choice experiment itself was designed to resemble competing bids for network management systems from two different vendors. Examples of these proposals from competing vendors are shown in Figures 3 and 4. Each alternative NMS product was described in terms of eight attributes which include the vendor offering the system, the interface provided to other vendors' equipment, several different types of functionalities, and price. Although it cannot be seen on these figures, the software is designed to highlight through shading any of the attributes which differ between the two proposals. The manufacturers' network elements for which different types of functionality or connectivity could be provided were based on the information provided by the respondent about the corporation's network elements.

A price algorithm was programmed into the software to price out each network management system configuration based on the respondent's existing network characteristics and the levels of functionality offered. The price algorithm consisted of many different price components, each of which was assigned one of three possible levels. For each of the two proposals being compared, the same levels were used for the same price components. The price of a system shown in each proposal consisted of a one-time charge and a monthly recurring charge.

For each pair of proposals, the respondent was asked to express a preference between the two and to then rate the likelihood of purchase on a scale of 1 to 10 for the chosen proposal. Ten of these pairs were presented to each respondent. As discussed earlier, the preferences expressed in a survey setting tend to be more relevant to actual market behavior when the respondent is well motivated and has a high level of involvement in the choice tasks. It was advantageous, therefore, to be able to use the software's internal timer to record the length of time spent on the survey by each respondent. A small amount of time spent on the survey was an indication that the respondent had not become highly involved in the choice task; these diskettes were discarded due to the possible invalidity of the survey responses.

In general, this study serves as another example of a successful disk-by-mail survey application. A response rate of 55 percent was achieved and respondents were quite enthusiastic about the approach. In particular, this application demonstrates the feasibility of using a disk-by-mail survey for an extremely complex and innovative stated preference experiment.

NETWORK MANAGEMENT SYSTEMS — P. 1 Q. 1

Good afternoon, Mr. Parker!

Thank you very much for agreeing to participate in our study of network management systems (NMS). On this diskette we will ask you questions regarding the following:

- The types of network components and applications that you would want to include as part of an NMS; and

- Your preferences regarding various features and functions associated with such a system;

Press <Esc> to go back or any other key to continue

FIGURE 2.

EQUIPMENT TO BE INTEGRATED — P. 2 Q. 4

First, we would like to find out which specific types and vendors of equipment you would include as part of a network management system (NMS), if you were to purchase such a system in the coming year.

Would you include MODEMS?	Yes
What is your primary vendor for MODEMS?	AT&T
Would you include a second vendor's MODEMS?	

	4 IBM	7 Racal-Vadic
2 Codex	5 Paradyne	8 Other (specify)
3 Gen.DataComm	6 Racal-Milgo	N No others

Please give your answer above (or <Esc> to go back one screen)

FIGURE 3.

Which of the NMS product proposals shown below would you prefer most?

Proposal A from IBM		1	Proposal B from AT&T	
VENDORS	FEATURES		FEATURES	VENDORS
-Primary-	>>> FAULT FUNCTIONALITY		<<<	-Primary-
IBM/SNA host	Level 5		Level 5	AT&T modems
	CONFIGURATION FUNCTIONALITY			
	Fully Automatic		Manual	
	OTHER FUNCTIONALITY			
	Performance Mgmt.		Security Mgmt.	
-Others-	>>> NON-SNA INTERFACE		<<<	-Others-
AT&T modems	Terminal Emulation		None	No.Telecom PBXs
No.Telecom PBXs				MCI T-carrier
and others				
SNA Network	>>> SNA INTERFACE		<<<	SNA Network
IBM	Full Two-Way Comm.		Alarms to SNA	IBM
	SNA MANAGEMENT			
	Entire Network		Backbone Network	
	PRICE AND CHARGES			
	$ xxxx plus		$ zzzz plus	
	$ yyyy/month		$ wwww/month	
Prefer System A			Prefer System B	

Use the cursor arrows and <Return> to select (or <Esc> to go back one screen)

FIGURE 4.

How likely is it that your company would purchase SYSTEM B?

Proposal A from IBM		1	Proposal B from AT&T	
VENDORS	FEATURES		FEATURES	VENDORS
-Primary-	>>> FAULT FUNCTIONALITY		<<<	-Primary-
IBM/SNA host	Level 5		Level 5	AT&T modems
	CONFIGURATION FUNCTIONALITY			
	Fully Automatic		Manual	
	OTHER FUNCTIONALITY			
	Performance Mgmt.		Security Mgmt.	
-Others-	>>> NON-SNA INTERFACE		<<<	-Others-
AT&T modems	Terminal Emulation		None	No.Telecom PBXs
No.Telecom PBXs				MCI T-carrier
and others				
SNA Network	>>> SNA INTERFACE		<<<	SNA Network
IBM	Full Two-Way Comm.		Alarms to SNA	IBM
	SNA MANAGEMENT			
	Entire Network		Backbone Network	
	PRICE AND CHARGES			
	$ xxxx plus		$ zzzz plus	
	$ yyyy/month		$ wwww/month	

Definitely NOT < 0 1 2 3 4 5 6 7 8 9 10 > Definitely BUY

Give your answer on the scale of 0 to 10 (or <Esc> to go back one screen)

MAINTAINING QUALITY IN LARGE COMPUTER-INTERACTIVE INTERVIEWING PROJECTS

Nancy L. Messinger
Burke Marketing Research

While computer-interactive interviewing is a relatively young innovation for market research, Burke Marketing Research is a relatively old company. Founded in the 1930's, it has had more than 50 years to develop standards for the execution of traditional paper and telephone interviews. As the company grew, it compartmentalized research activities into specific job functions. Internal bureaucracies of considerable size developed to support these job functions. The execution of studies became the responsibility of a large department that over the years developed a specialized field subcontracting office. Data aggregation became the responsibility of a report processing department. Data reporting (under clear-cut bureaucratic strictures) occasionally was completed by an analytical department. Quality control was managed by a project services department, responsible to itself initially, but more recently reporting to client services. Burke had grown internally to the extent that the organization was geared to handling the very largest studies with the highest attention to quality and detail, but only doing so in the context of data acquisition via paper or telephone.

When computer-interactive interviewing on personal computers began to take off about four years ago, it was introduced to Burke by a few brave innovators who believed that it was the wave of the future. Needless to say, a large bureaucracy built for gathering data through other media was a little perplexed, and at times resistant to this new technology. Over time, the innovators managed to execute a few studies and demonstrate the virtues of a computer-assisted data collecting system. At present, Burke is in a position where computer interactive interviewing has been adopted. We have determined the likely share of our business that will involve computer-interactive interviewing and are presently introducing quality standards for the various departments involved in execution, tabulation, and reporting of data collected from computer-assisted interviewing.

As mentioned earlier, we are in the process of modifying procedures that were incorporated by various departments during the initial adoption process. Based on the quality control and efficiencies within the entire framework of the company, some of the departmental procedures have proven to be cumbersome. Decisions that have proven to strengthen the overall structure of the company have been documented and are being used to form the foundation for our computer interviewing standards. We have come to realize that there are four basic components to the successful execution of a large computer-interactive study:

- Our own efforts in orchestrating the study.
- The capabilities of the interviewing software.
- The respondents' capabilities to perform during the interview.
- The field agencies' ability to execute the study.

I've been asked to discuss a few of the details of maintaining quality in large computer-interactive interview studies. Some of the procedures I would like to share with you are currently in a developmental stage at Burke, some techniques have not worked well for us, and others have provided results better than we could have anticipated. The specific areas I will address
include:

- KEEPING TRACK OF DETAILS:

  COMMUNICATION

  DISK NUMBERING PROCEDURES

- ADVANTAGES OF COMPUTER INTERVIEWING

- DISADVANTAGES OF COMPUTER INTERVIEWING

- CAUSES OF COMPUTER-PHOBIA

  OVERCOMING COMPUTER-PHOBIA: INTERVIEWERS

  OVERCOMING COMPUTER-PHOBIA: RESPONDENTS

- COMPUTER INTERVIEWING COSTS

- COMMON DATA ACCUMULATION PROBLEMS

KEEPING TRACK OF DETAILS: COMMUNICATION

The larger the organization, the more critical communication becomes in assuring that quality is maintained throughout the entire data processing system. Burke, like many other marketing research organizations, has sales offices across the country staffed with account executives. This sales force is responsible for working with clients to assure specific study objectives are well-defined and addressed through the marketing research process. Project directors, primarily located in Cincinnati, monitor the costs associated with a job and oversee the daily execution of studies. Analysts, also located in Cincinnati, are responsible for the initial statistical design of studies and the interpretation of the collected data. Though these are usually the employees that are visually associated with a study, there are other individuals that must be included in the communication process for quality to be maintained throughout data collection and processing.

For large studies involving computer interactive interviewing, Burke has found a "Who's who" document to greatly improve the flow of information within the company. The names of ALL individuals associated with a specific study are recorded in this document including the account executive, project director, analyst, computer programmer, those who contract with field agencies, data coding and entry personnel, and data processing staff.

Phone numbers, both home and office, are listed for each individual. Business addresses or internal mail locations are also provided for each member. A brief paragraph is written for each person, detailing specific job responsibilities. Included with these responsibilities are information on when resources will become available to begin work on a study, and when specific tasks should be completed. Another important component of this document is the name and phone number of the person (or people) that should be contacted if ANY problems should arise during the study.

A copy of the "Who's who" document is given to each person involved in the study. This type of document has greatly helped to improve the communication process by ensuring that each individual's responsibilities are clearly understood by all members involved with a specific study.

KEEPING TRACK OF DETAILS: DISK NUMBERING PROCEDURES

Careful consideration should be given to the designation of respondent numbering schemes for rapid identification of field locations and to maximize usefulness for subsequent data segmentation.

For example, the six-digit code best suited for a specific study may consist of the following format:

#	X X	X	X	X X
	City	City Location	Product Region	Respondent Number

Where:	City	=	10 to 17 (8)
	Location	=	1 to 4
	Region	=	1 to 4
	Resp. Number	=	1 to 99

Most of the computer interactive interviewing studies conducted at Burke have involved more than data collected via computers. After the data has been collected from computer interviews, the merged data set is often loaded into a data management program (Lotus 1-2-3, SAS, SPSS, etc.) and merged with data collected from other methods of interviewing. To date, the advantages of maintaining internal consistencies across all computer- interactive interviewing studies have greatly helped to maintain quality standards at Burke. In our example, if eight cities are being used to field this study, the cities would be numbered from 10 to 17. The study is being conducted in one to four locations in each of the eight cities, and is designated in the third column. Our client has products that are located in four distinct regions of the country, which are noted in column four. The last two digits in the coding scheme are reserved for the specific respondent number.

Included in the "Who's who" document is a "Master List of Interviews and Disks." (Figure 1). The list follows the numbering scheme described above, and is divided into two sections. The first, "Required Disks," contains information on the disks that are required to complete the quota in a specific location. The second, "Extra Disks," describes the extra disks that have been sent to each location. In general, a minimum of 10% extra interview space is sent to each field location.

The "Master List" is interpreted as follows:

In Denver, two different locations are fielding the study.

Denver has client products from the fourth product region.

Forty interviews are required from each location.

If five interviews are allotted per disk, eight disks will be required to meet the quota.

For city three, location one, required respondent numbers range from 1214<u>01</u> to 1214<u>40</u> (5 * 8 = 40 interviews).

The two extra disks sent to city three, location one had numbers from 121441 to 121450.

One person is assigned the responsibility of labeling the disks with appropriate respondent numbers, mailing the disks to the field locations, and keeping track of numbering for any subsequent disks that are needed in the field.

By using the six-digit respondent field in this manner, data is very easy to subdivide at a later date based on pre-determined categories, and any individual associated with this study will be able to quickly identify the city, location within a city, or the product region from which the data was collected.

ADVANTAGES OF COMPUTER INTERVIEWING

The advantages of using computers to collect survey data have been fairly well documented. The three basic advantages for Burke include:

- <u>Reduced Interviewer Bias</u>
  Minimal interaction of the interviewer with the respondent will reduce or eliminate many forms of verbal and non-verbal interviewer bias. (However, new forms of bias may be introduced.)

- <u>Execution of More Complicated Analytical Packages</u>
  Various analytical techniques such as conjoint analysis can be performed more quickly and thoroughly with computers than with conventional pencil-and-paper methods.

- <u>"Minimizes" Data Entry Errors</u>
  Assuming ALL the data are collected by entering the responses directly into the computer, and that no data entry errors are made from the respondent mis-entering the data into the computer, errors from data entry can potentially be eliminated.

DISADVANTAGES OF COMPUTER INTERVIEWING

Since most of the computer interactive interviewing studies conducted by Burke in the past have required more than the computer alone to collect the data, data entry problems can still exist. When paper screeners are used, or paper sections of the questionnaire, or lengthy open-ended questions are asked using computer interviewing, coding of the responses, data entry, and merging of the paper portions of the study back into the computer portions require human intervention with the data set. In these situations, a new and different form of data entry bias is encountered. In the best situation, data would be collected using only a computer data-entry process. So far, very few studies Burke has encountered have allowed us this luxury.

In the past, open-ended responses have presented one of the most difficult data collecting problems using computer-interactive interviewing. Respondents have had the option of typing their own responses, writing their responses on a sheet of paper, or having the interviewer enter responses as dictated by the respondent. Each of these methods has its inherent problems and biases.

In a recent computer study where open-ended questions were asked and respondents were asked to type their own answers, several unexpected problems occurred: After each line of text was entered by the respondent, the "enter" key had to be pressed to move to the subsequent line on the computer screen. Many typing errors were made having to back-space over incomplete words before pressing "enter."

Another problem recently encountered was the number of lines on the computer screen available for lengthy open-ended responses. As long as the answers were brief, space was not a problem. If answers required more than one screen, the text became tedious to input and very difficult to edit.

The typing ability of the respondent played a major role in the "successful" entry of text for open-ended questions. With few exceptions, we have found the typing skills of the respondents to be fairly poor. With the variety of problems that have been encountered by allowing participants to enter their own responses, other methods of recording open-ended responses are being explored.

CAUSES OF COMPUTER PHOBIA

From our experiences, computer phobia is a very serious problem that both respondents and field interviews suffer to varying degrees. The primary cause of computer phobia appears to be quite simple: Minimal past involvement with computers. This condition may include from not knowing how to turn on a computer, poor keyboard skills, a lack of knowledge of computer terminology or functions.

When conducting a computer-interactive interview, it is very important that field facilities have a basic level of computer literacy. Common situations/problems that are encountered when fielding a computer study include:

- Rental computer equipment problems.
- DOS incompatibility problems\outdated DOS versions.
- Disk density incompatibilities.
- Disks that will not boot.
- Disks damaged in shipping.
- Delays in shipping/lost disks.
- Reformatting disks/erasing data.

Another problem Burke has encountered in several instances involves the interviewers knowing just enough about computers to inadvertently cause problems with the data. For example, in trying to read a data file on a disk to assure that all data has been appropriately collected, data files have been erased/damaged and entire disks have been reformatted. (A possible solution to this problem is presented later in this paper.)

OVERCOMING COMPUTER PHOBIA: INTERVIEWERS

When problems are encountered in the field while conducting computer interactive interviews, there are several services that marketing research firms can provide to minimize field problems and maintain quality while fielding studies.

- Provide a Standard Operating Procedures (SOP) Manual
  The SOP manual should provide highly detailed instructions for conducting computer-assisted interviews. A manual should be kept by each field location as a quick reference for most problems that are encountered.

- Provide Job-Specific Instructions
  For each computer study that goes to the field, a special set of instructions should be included with the regular field materials. These instructions will provide any additional instructions that are needed in the execution of a specific computer-interactive interviewing study.

  Included in these special instructions should be the names and phone numbers of the individuals that should be contacted should any problems arise that are not addressed in the SOP or Job-Specific Instructions. In addition, the times of day when each person can be contacted for assistance is extremely helpful.

OVERCOMING COMPUTER PHOBIA: RESPONDENTS

There are several steps that can be taken to minimize the severity of computer phobia for respondents.

- Keep Respondent Instructions Simple
  Keep the initial instructions in the interview very brief and easy to follow. Use several screens to explain the instructions rather than cluttering one or two screens. Color can be useful in separating sections of instructions, highlighting important points, or simply improving the readability of the computer screen.

- Start with Non-Threatening Questions
  Start with very simple questions, possibly binomial in design, to give the respondent a chance to experience the keyboard, relax, and build confidence. Slowly move into the more complex types of questions.

- Have an Interviewer Nearby at ALL Times
  Have an interviewer in the room at all times in a non-imposing location. If a problem arises, someone should be nearby to provide immediate assistance without crowding, intimidating, or in any way restricting the respondent.

COMPUTER INTERVIEWING COSTS

There are several costs associated with execution of a computer- assisted interview study that are inherently different from other more traditional methods of data collection. During the development of a study, there are internal costs (I) associated with computer studies before any data is collected. Once the disks have reached the field, there are external costs (II) charged to the job by the field agencies for collection of the data. After data collection, another series of internal costs (III) is incurred before the data are ready for interpretation.

I. Initial Internal Costs

- Programming Costs
  The equipment depreciation and employee time costs associated with the programming of a study.

- Cost of the Diskettes

- Formatting/Copying Costs
  Labor/equipment costs for formatting disks, copying the program to the disks, and using the label program to number and identify each of the disks.

- Quality Check
  After the disks have been placed back in their boxes (10 per box), one disk is removed from each of the boxes and checked to make sure it is working properly. If a disk is "bad," the entire box of disks where the bad disk was found is checked. At least 10% of the disks are always checked, but the percentage of disks that are evaluated is dependent upon the requirements of the study.

- Shipping Costs
  The cost of packaging and mailing the computer disks to the various field locations.

II. External Field Agency Costs

- Recruiting Costs
  Including respondent incentives.

- Interviewing Costs
  Generally around $20.00 per hour.

- Computer Equipment Costs
  If field agencies own their own PC's, a nominal fee is charged for the use of their equipment. If the agencies have to rent PC's for a study, the cost of the rental is passed on to the marketing research firm, often with little or no mark-up from the agency. Rental fees for a single computer seem to range from $100.00 to more than $200.00 per day.

A Comparison of External Costs

A number of past studies conducted by Burke were reviewed to determine the Cost Per Interview (CPI) for a study conducted without using computer-assisted interviewing, for a study using computers that were owned by the field agencies, and for studies where PC's had to be rented to conduct the interviews. The data collected for each field location included the total cost of the study, the final number of completes, the average length of the interview, and the total number of contacts made in recruiting for the study.

The effective incidence level for each location was calculated by dividing the total number of contacts made by the total number of completed interviews. The CPI was calculated by dividing the total cost from the field location by the total number of completes. Since interview length would have an effect on the cost per interview, the CPI was divided by the average number of minutes required to complete the questionnaire.

In Figure 2, the raw data points have been plotted to show the relationship of not using a computer to collect data to using a rented PC's or a PC owned by a field agency. The effective incidence levels evaluated in this exercise ranged from 2% to 38%. The average CPI per minute for conducting a study that does not use computer-assisted interviewing is approximately $1.39 (Figure 3). When computers are used in the study to collect data, the average CPI per minute increases to $1.54. However, when the average CPI per minute of studies using a computer owned by an agency are compared to studies that include the cost of computer rentals, most of this increase in cost is attributed to the incremental increase of agencies having to rent PC's to conduct studies (Figure 4). The average CPI per minute to conduct a study using PC's owned by agencies is $1.37; almost the same as not having used computers in the study to collect data. The average CPI per minute for studies where agencies have to rent computer equipment is substantially higher ($2.05).

Controlling External Costs

Based on the above findings, the easiest way to reduce the cost of data collection in the field is to limit the number of studies being conducted with rented PC's. If PC's need to be rented for a study, limit the number of machines that an agency has to rent. With most studies, field agencies rent two machines to allow for simultaneous interviewing. In the past, Burke has asked agencies to rent a "back-up" PC in case one of the PC's is not operating properly. We are currently evaluating a system to eliminate the need for renting a back-up computer, while ensuring that the data being collected are sound.

Process to Eliminate "Back-Up" Computer Rentals

Two days before a study begins in the field, practice disks will be sent to each mall location. One practice disk will be sent for each computer being used for the study. The day before the beginning of the study, the field agencies would rent the standard two computers and run through the practice disks. When completed, the disks would be sent back to Burke using an overnight delivery system at a cost of not more than $7.00 per location. By 10:00 am the day of the study, the disks would be delivered to Burke. The data files would be checked to make sure all equipment is working properly. The field agencies would be telephoned by noon to either alert them of data problems or to tell them to begin the interviewing process.

By using this system to check the data, Burke would be assured that all equipment was working properly and that the interviewers knew how to operate the computers properly without having to incur additional costly rental charges.

III. Final Internal Computer Costs

- Return Mailing of the Disks to Burke

- Accumulation and Merging of the Data

- Final Restructuring of Data
  The cost involved with manipulating the data to run cross-tab tables or other interpretations of the data.

COMMON DATA ACCUMULATION PROBLEMS

The most common problems encountered while accumulating data from computer-interactive interviewing studies include the following:

- Duplicate Respondent Numbers
  This problem has occurred with pre-numbered and self numbered interviews. (In some cases, we have not been able to replicate the errors observed.)

- Unidentified Respondent Numbers
  The occurrence of respondent numbers that do not follow a numbering sequence or have any apparent identifying features.

- Missing Data
  Numbering sequences with missing members.

- Damaged Disks
  On occasion, disks are damaged by faulty computer equipment, by the careless handling of disks, or by damage incurred during the shipping process.

- Blank Disks
  Sometime during the fielding or mailing process, disks are either reformatted or completely erased.

MODEMS

In the near future, one method of eliminating many of the problems encountered when trying to conduct large studies using dozens of disks may be the increased use of modems for data transfer. With more and more field agencies buying their own computers and modems, the reduction in disk handling activities may be the most reliable and cost effective method for executing quality computer-interactive interviewing studies.

SUMMARY

Burke is currently in a period of transition, and in many ways it is fortunate that computer-interactive interviewing is being adopted in the current environment. Reorganization of our business for the 1990's allows the perfect climate for learning how to better incorporate new technologies into our existing framework. Computer-interactive interviewing has progressed well beyond the stage where it is simply an innovation. As it becomes a larger part of our business, we will demand that many of the problems that we encounter be solved, not only internally, but externally as well.

The key to improving existing processes and adopting new technologies is communication. When problems are encountered anywhere within the data collection and processing systems, they must be evaluated, discussed, and resolved. Our ability to address these problems openly, in concert with all concerned parties, will ultimately determine the quality of large computer- interactive interviewing studies.

FIGURE 1

MASTER LIST OF INTERVIEWS AND DISKS

			REQUIRED DISKS			EXTRA DISKS	
CITY	LOCATION	REGION	INTERVIEWS PER LOCATION	NUMBER OF DISKS	INTERVIEW NUMBERS	NUMBER OF DISKS	INTERVIEW NUMBERS
BIRMINGHAM	1	2	25	5	101201-101225	2	101226-101235
CHICAGO	1	1	31	7	111101-111135	2	111136-111145
"	2	1	31	7	112101-112135	2	112136-112145
"	3	1	31	7	113101-113135	2	113136-113145
"	4	1	31	7	114101-114135	2	114136-114145
DENVER	1	4	40	8	121401-121440	2	121441-121450
"	2	4	40	8	122401-122440	2	122441-122450
HOUSTON	1	3	41	9	131301-131345	3	131346-131360
"	2	3	41	9	132301-132345	3	132346-132360
KANSAS CITY	1	1	31	7	141101-141135	2	141136-141145
"	2	1	31	7	142101-142135	2	142136-142145
LOS ANGELES	1	4	38	8	151401-151440	2	151441-151450
"	2	4	38	8	152401-152440	2	152441-152450
"	3	4	38	8	153401-153440	2	153441-153450
MIAMI	1	3	27	6	161301-161330	2	161331-161340
"	2	3	27	6	162301-162330	2	162331-162340
NEW YORK	1	2	24	5	171201-171225	2	171226-171235
"	2	2	24	5	172201-172225	2	172226-172235
"	3	2	24	5	173201-173225	2	173226-173235
"	4	2	24	5	174201-174225	2	174226-173235

FIGURE 2

INTERVIEWING COSTS

NO COMPUTER VS. COMPUTER (RENT/OWN)

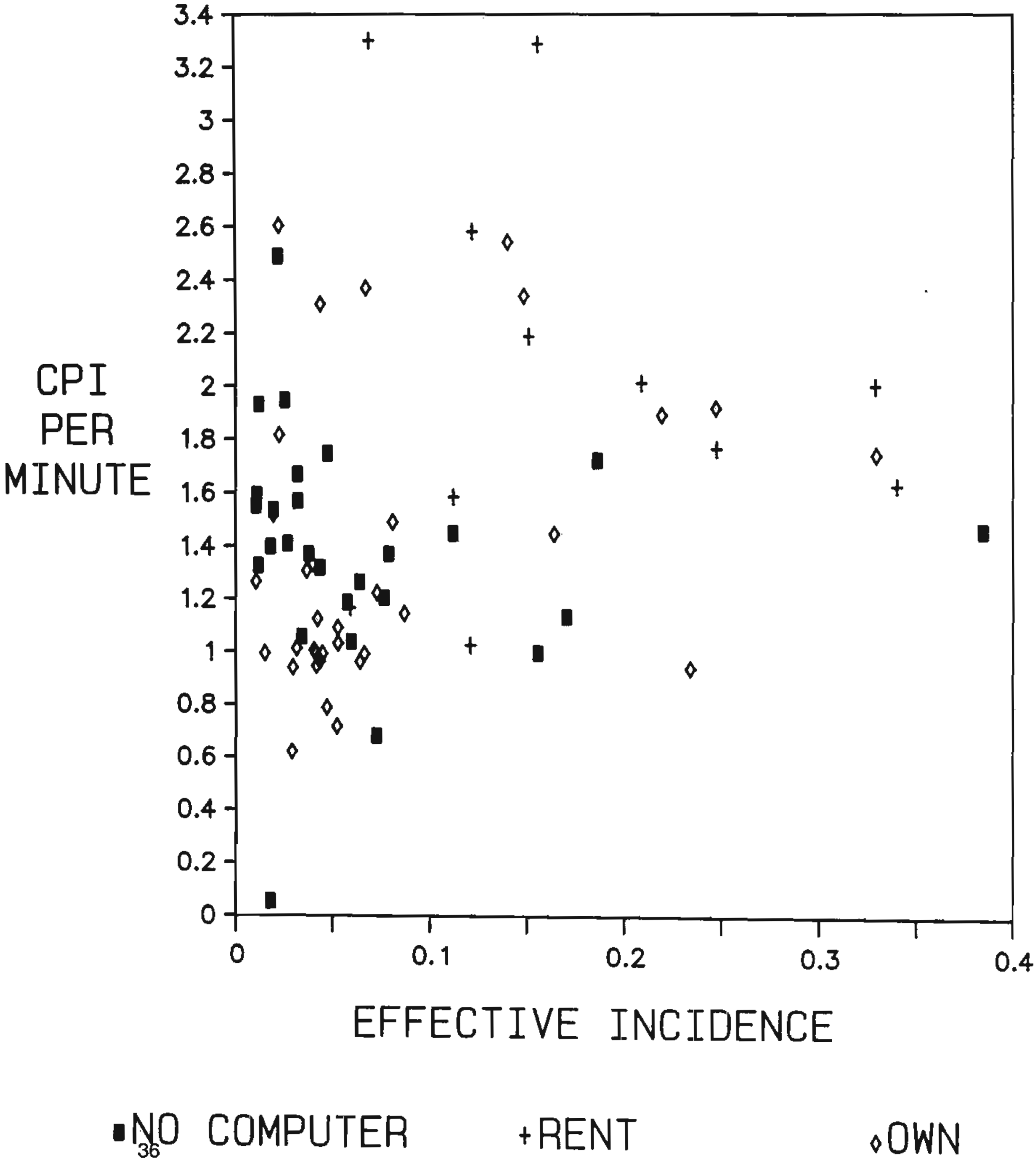


FIGURE 3

INTERVIEW COSTS

NO COMPUTER VS. COMPUTER

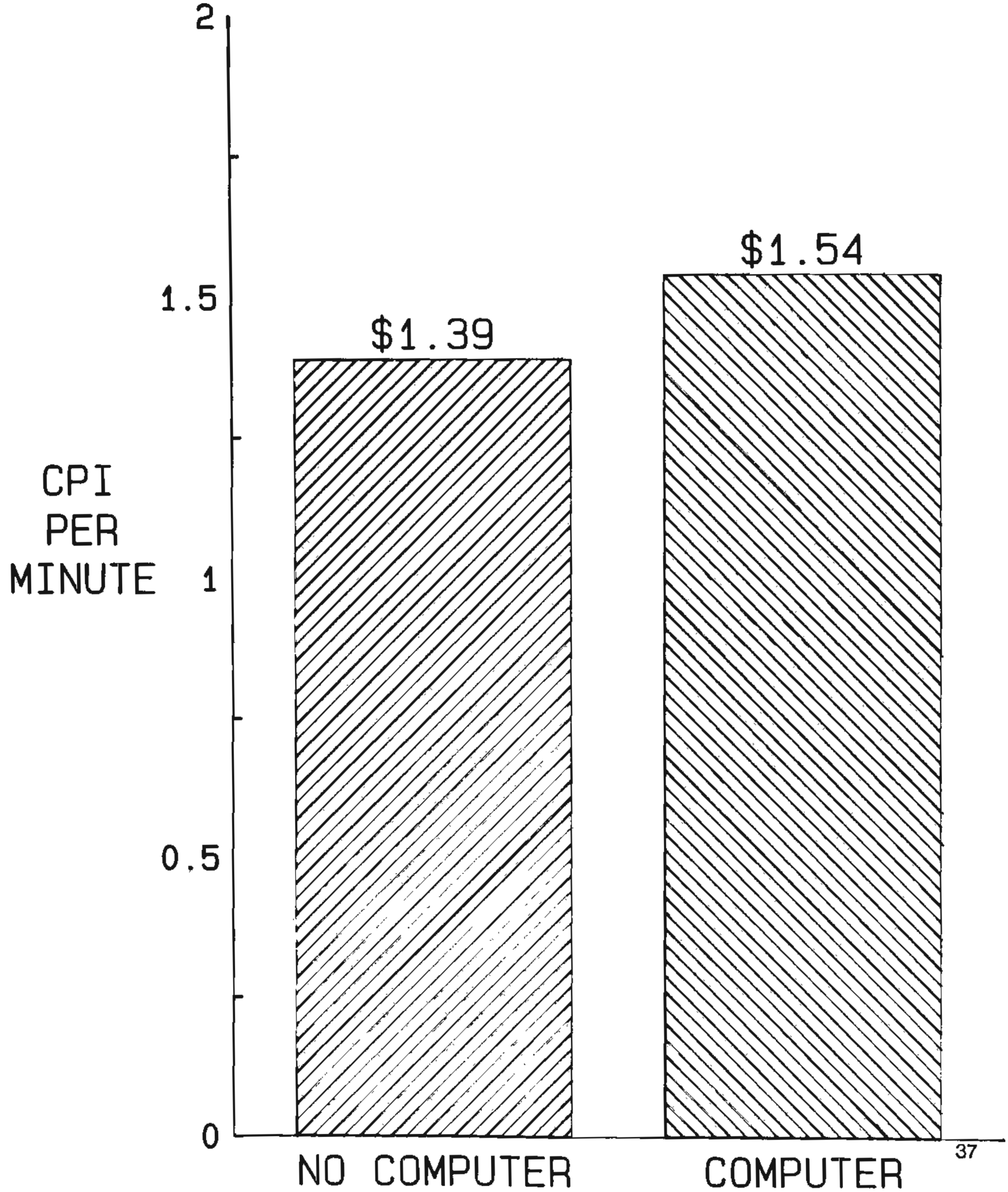
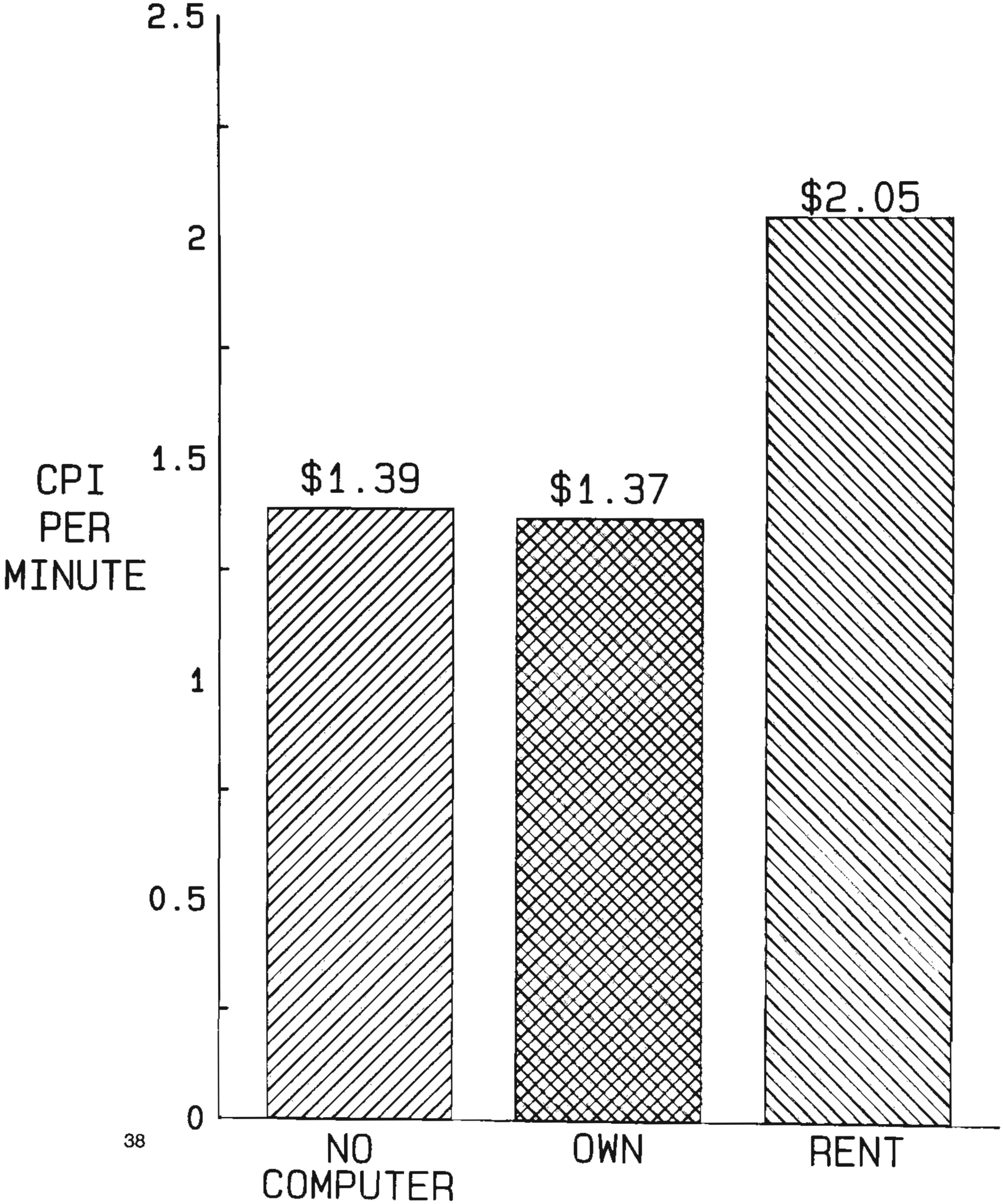


FIGURE 4

INTERVIEW COSTS

NO COMPUTER VS. COMPUTER (OWN/RENT)



20,000 COMPUTER INTERVIEWS A YEAR: IS THIS INSANITY?

Diane L. Pyle
Hallmark Cards, Inc.

This paper explains how the Marketing Research Division at Hallmark Cards, Inc. uses PC interviewing with a particular methodology in administering more than 20,000 interviews a year. The content of the paper will be as follows:

- a description of the background of the process,
- a description of how it was previously done with pencil and paper,
- the decision to convert it to PC,
- the process as it is done today on PC, and
- the differences between the pencil-and-paper version and PC-version

And in the process of explaining this very laborious transition, I will most generously reveal my secrets as to how to maintain sanity.

BACKGROUND

At Hallmark one of the questions that Product Management needs answered time after time is, "What items should we be offering our customers?" One of the ways Hallmark Marketing Research answered this question is by using an "Item Screen" questionnaire. The questionnaire has been used successfully for a number of years for selecting the best items to go into our retail outlets. In 1985, with the upcoming launch of the Shoebox Greetings ("a tiny little division of Hallmark Cards") card line, there was a need for 400-500 new cards. Three items had to be tested for each one item to be included in the product launch. That meant over 1500 cards needed to be evaluated by consumers! The "Item Screen" was chosen as the method for evaluation.

An "Item Screen" is just what the name implies: It's a certain number of items that consumers evaluate or "screen" on a six-point scale (excellent to poor). Consumers indicate what items they are most likely to purchase and demographic information is collected. The questionnaires are not "canned" by any means; there is a great amount of flexibility in the up-front screening and add-on questions that may be asked. Still, the basic body of the questionnaire remains constant.

The results of the testing proved successful in the launch of the Shoebox Greetings card line and, thus, continuous Item Screens were used to keep the line fresh with the addition of new cards. Hallmark obviously wants their stores to stock those items that their customers will want to purchase now. "Now" implies getting the cards chosen by the Item Screens manufactured quickly; time becomes a crucial element. Because of the time factor and the volume of Shoebox Item Screens needed to keep the line fresh and consumer driven, the administration was standardized. Standardizing the process guaranteed quick turnaround and lowered costs. The elements in the standardization included:

- 50 designs to be tested in each group,
- prototype cards mounted on cardboard and numbered with the same set of 50 numbers used for each study,
- cards grouped for rotation and rotated after every 25th respondent,
- each test being sent to two cities with 100 respondents per city,
- schedules dictating when the test material is to be delivered to Research, how long the study will be in the field, and when final reports are due to product people, and
- the questionnaire being self administered and taking approximately 20 minutes to complete.

PENCIL-AND-PAPER ADMINISTRATION

The Shoebox Greeting cards were delivered to Marketing Research by Hallmark Product Management. Marketing Research contracted with a field service to administer the study. The selection of the field service had some constraints because of the amount of space needed to conduct the study. The study had to be set up in only one room and had to accommodate four to five 6-foot tables or 24-30 feet of tabletop space. The tables were sectioned off into ten sections, each section containing a group of five cards.

A respondent who passed the screening requirements and agreed to participate was taken to the room and given the questionnaire to complete. The respondent would then walk beside each table, evaluating each card as it was laid out on the table.

The questionnaire instructed the respondent to write the number of the card in the appropriate space and then indicate the evaluation for each of the 50 designs being tested. The respondent then chose up to eight cards he or she would most likely purchase out of the group of 50, recorded the card numbers, and indicated the relationship of the person the card would be sent to and the approximate age of the recipient. A human monitor was in the room at all times to make sure respondents followed the specific instructions and to check the questionnaire for completeness before the respondent left. When a test location had completed 100 interviews, questionnaires and testing materials were sent back to Hallmark. The questionnaires were edited by Hallmark employees to ensure that the respondent had correctly entered card numbers, had evaluated all 50 cards, and had no double evaluations for any cards. This process took approximately 8 hours per study.

Good questionnaires (approximately 90-95%) were sent to the data input department for coding; questionable questionnaires were given back to the analyst in charge of the study to decide whether they were to be used or discarded. Coding, keypunching, and verification took a day and a half to complete. The reports were then run and given to the analyst.

All in all, from the time that the field people received the questionnaires and testing materials until the time the analyst had the reports in hand took approximately 35 days.

VOLUME INCREASES

In 1985 approximately 25 studies (5,000 interviews) were done with 50 items being evaluated. These were nonstandardized screens used in the initial launch of the Shoebox Greetings line. Forty-seven studies (9,400 interviews) were done in 1986; this included the addition of a second product line for which management needed new stock numbers.

The testing process had proven itself very successful, especially in selecting humor items that retailed well. Hallmark has a good strong foundation within its Creative ranks. They recognize what is funny and are able to create it. However, Product Management finds it hard to understand what humor is appropriate for specific sending situations. That is, even though something is funny, is it still funny when sent to relay a message, express gratitude for something special someone's done, or perhaps just to brighten someone's day? Item Screening was able to determine this and other humor-driven lines were added.

The next two years showed a dramatic increase in tests: 85 studies (17,000 interviews) in 1987 and 117 studies (23,400 interviews) in 1988.

DECISION TO CONVERT TO PC VERSION

In 1988, the volume of the tests had reached a level which demanded some changes. There was at least one test every week, and in many instances, three to four a week. Because of their importance in planning product lines, these studies took priority in the data processing area. Since these studies came first, everything else was secondary and delays often occurred with the processing of other studies. Though it may be extreme to say insanity was running rampant, a high level of stress was created for analysts waiting for study results and for personnel in data processing with unreasonable workloads. Either more employees had to be added or an alternative to the current Item Screen system needed to be implemented.

Two options were considered: one was having the studies contracted to an outside supplier; the other was to convert the testing to PC and eliminate in-house editing, coding, and keypunching.

So, the first thing we had to do in the evaluation phase was to convert the Item Screen Questionnaire to a PC version. (The actual structure of the PC questionnaire will be discussed later in the paper.)

The second step was to determine if the PC interview should be interviewer or respondent administered. The average time for a respondent to complete a questionnaire was 28.6 minutes; the average time for the interviewer to complete it was 18 minutes. Because of the tight turnaround time required on these tests and the control needed, Hallmark decided the interviewer should administer each questionnaire.

Three studies were run using the pencil-and-paper method and then they were re-fielded and run with the PC method. The two sets of data were compared against one another for any variances. The comparisons showed very few differences between the two methodologies. The mean scores were similar and rankings of the top 2 boxes had high correlations. The only difference was that the raw percent in the top 2 boxes was somewhat lower in PC tests. (The reason for this will be explained later.)

Of the two options being considered - either contracting to outside suppliers or staying in-house and converting to PC data gathering - it was determined more cost effective to keep the Item Screening in-house and convert to PC. The conversion to PC administration of Item Screen studies took place in August, 1988. Twenty-one studies (4,200 interviews) were completed using the PC in 1988, and 1989 found 105 studies scheduled, including 16 studies for two additional product lines. Over 20,000 interviews are expected to be completed in 1989!

THE ITEM SCREEN PROCESS ADMINISTERED USING THE PC

The up-front work on an Item Screen study using a PC is very similar to that used with the pencil-and-paper methodology. The study is received in Research and a field service is contracted with to administer the study. One advantage in fielding a PC study is that display space is no longer a constraint since the PC administration does not necessitate the displaying of the cards on a table. The only requirements are a facility that has two PC's and a staff that can work well with Hallmark.

In the packet of information sent to the supplier is a Ci2 (Ci2 System for Computer-Interactive Interviewing) "Personal Computer Interviewing Tips" document. The document includes information on DOS (Disk Operating System) commands, Ci2 commands, and information based on previous field services' experiences with PC studies. We also send the field a form to document the number of interviews completed on each disk and to record any comments on a per-disk basis. These documents help in the control of the study.

Preparing field disks soon proved to be too time consuming, so master disks were sent to an outside disk duplicating firm. Because of preset schedules for all Item Screen studies, we know the number of disks we'll need for the year. The company we use is Transmedia Technology, 10111 Santa Fe Drive, Overland Park, KS 66212. Transmedia provides the duplicated disk, disk sleeve, and labeling. The only step left to be done in-house is running Ci2's SETNUM program to pre-number the respondents.

Though the questionnaires remain primarily constant, sometimes product line managers will decide to add an extra question or make some other change. Transmedia's policy is to charge half the original duplication cost when re-duplicating disks.

Disks are shipped in protective, magnetic resistant disk mailers which hold five disks. We allow 25 respondents per disk, so five disks are mailed for each study. Once the Item Screen is in the field and a respondent has passed screening requirements and agreed to participate, the respondent is taken to the location of the test and asked to sit next to the computer. The cards are stacked next to the computer in groups of five. The five cards for each of ten groups are determined prior to being sent to the field service and cards remain in their respective group throughout the study. The cards are naturally rotated within their group as the respondents go through them. A rotation pattern is given to the field service personnel, and they rotate the cards every 25th respondent.

The interviewer hands the respondent each group of five cards in the proper rotation pattern.

Initially, the respondent is instructed to hold out those cards he or she would most likely purchase while evaluating the cards. The respondent reads the card number to the interviewer sitting at the computer. The interviewer inputs the number and then records the respondent's evaluation. If any number is input that is not a valid card number, the interviewer receives a message from the PC. Should a card number be given that has already been evaluated, the logic of the questionnaire takes the interviewer to a screen which gives two options: 1) to return to the input screen (an input error has occurred), or 2) to re-evaluate the card (if the respondent changes an evaluation on a particular card).

An internal counter is kept to ensure that all 50 cards are evaluated. Should the interviewer try to go ahead with the interview without completing the evaluation of all 50 cards, a message appears on the screen listing the cards evaluated and not evaluated. The interview can not continue until all 50 cards are evaluated.

The PC questionnaire "pre-edits" the respondent's data. Previously, questionnaires were edited manually, after interviewing, to ensure a respondent had entered card numbers correctly, had evaluated all 50 cards, and had no double evaluations. The Ci2-programmed questionnaire guarantees these checks in the initial data entry.

Following the six-point scale evaluation of the cards, respondents are asked to indicate up to eight cards they'd most likely purchase. Since respondents have been holding their favorite cards out during the evaluation process, they are given time to select from those held out if they have kept more than eight. The favored card numbers are input one at a time. Again, the questionnaire is preprogrammed with the correct card numbers, so no incorrect card numbers can be input. Two questions regarding the age and relationship of the recipient are asked for each card. A few other questions are asked and then the respondent is asked to sit at the computer and answer some demographic questions.

In the pencil-and-paper administration of the questionnaire, the respondent can see all previous evaluations given. Therefore, he or she could be biased by previous evaluations and not want to rate everything too high or low. Also, with the cards laid out on a table, all cards are in view at all times. Potentially, the respondent could have been evaluating cards against one another, unintentionally. This is why the raw percent in the top 2 boxes was somewhat lower in the PC tests than in the studies administered using pencil and paper.

With the PC methodology, the respondent is geared to evaluating each card individually. As stated before, as each card is evaluated, it is either held out as a favorite or put on top of the stack of already evaluated cards. Also, unless respondents have memories as good as the computers recording their data, they aren't able to keep track of the scores they're giving each card. Therefore, in explaining the difference in the PC versus pencil- and-paper comparisons, the cards are kept close on means scores but the disbursement of the scores is changed because of the "individual evaluation" aspect which came into effect with PC administration.

The PC version offers a more true-to-shopping experience in the selection of the eight cards most likely to be purchased. Since respondents hold cards out as they initially go through them - instead of looking back over a table of 50 and choosing their favorites from that - it is more analogous to selecting cards in a card shop. Usually, card shoppers will retain those cards they're probably going to purchase as they make selections at a card display. Thus, not only is the PC interviewing method quicker, but it also allows us to capture more accurate data.

When a study is completed, the field staff is instructed to make a backup of the five disks before "air-expressing" them to Hallmark. Modem transfer of the data is being worked on and will soon be in place. Many facilities will not have this capability, so for the most part, the disks will continue to be air shipped.

With the large amount of data being returned to Hallmark Market Research, a natural question is, "How do we handle it?" Once the data disks are received, the data for each city are accumulated and converted separately. This is advantageous for several reasons. Disk can be accumulated as soon as they are returned without having to wait for a second location to be returned. Also, each city can be designated as a separate entity. This designation occurs by indicating a Job Number of 1 or 2 when converting the data with Ci2's conversion program. The cities' data are combined into a single file containing the complete study.

The accumulated and converted files for an individual study are copied to a single disk. This disk is designated for a certain study by having a Volume (disk) name given to it. This is accomplished by formatting the disk with DOS's FORMAT /V command.

The 80-column formatted file is then uploaded to the mainframe to a file with a name that designates the particular study. Reports are generated and delivered. Guaranteed backup is accomplished by the file being kept on the mainframe.

Hallmark Marketing Research has found the conversion of the Item Screen methodology to PC very successful. It has allowed human resources to be freed up (and their sanity maintained) to concentrate on unique projects, while the computer takes care of the routine, repetitive ones.

Several aspects allow those of us now involved with the PC Administration to stay sane. Those being:

- good communication with field services,
- organization - adhering to pre-set schedules and an organized system of processing data, and
- control - the number of people the process flows through is limited.

Of course, there are those days when things don't go exactly right. Like the one when we had data returned and one disk was completely blank - no data, no Ci2 files - NOTHING! After investigation, it was found that when the field service made backup copies of the disks before returning them, they mistakenly used the Format command (which completely erases a disk) for the particular disk in question instead of the Copy command. It's at that moment you wonder . . . *Is this insane?*

Optimal Pricing Strategies Through Conjoint Analysis

Mary Jane Tyner, Levi Strauss & Co.
Jonathan Weiner, MACRO

This paper focuses on the topic of obtaining better market intelligence and is concerned with optimal pricing strategies for a variety of jean styles.

Background:

Levi Strauss & Company, who produce a wide variety of men's jeans styles, finishes, and price points, were interested in measuring the price elasticity of their jeans. As a preliminary step, MACRO conducted an analysis of panel data collected for Levi Strauss & Company. After extensive regression modeling, a preliminary finding was developed that illustrated price elasticity for the LEVI'S brand, among different customer segments.

Before any decisions were made, management was interested in consumer feedback on price changes. It was also interested in knowing how various competitor responses to a LEVI'S price increase would affect sales and market share.

Research Objectives

The objectives of this research were:

- To determine the relationship between price and market share.

- To develop strategic pricing guidelines for LEVI'S men's jeans.

Research Methodology

A total of 304 men, between the ages of 15 and 44, were interviewed in shopping malls in 8 markets across the U.S.

Attribute and Level Specifications

Three primary attributes have been identified that impact product preference. They are jean brand, style, and price. These attributes are ideal for use in testing, because when the actual jean is shown to a respondent, the full jean profile, including brand and finish, is being presented to the consumer without any additional explanation.

Twelve brands and 5 styles were tested and all jeans were tested in one of five price categories. Relative to their average base price the jeans were priced at two levels above and below base as well as the average base price. The prices ranged from less than

$15.00 to slightly over $60.00. The combination of all attributes and levels yielded a total of 300 combinations. Using a one-sixth fractional factorial design reduced the number of combinations to be tested to 50.

Not all styles existed or were available for every brand, so the total number of jeans tested was 48.

Data Collection

The data collection was split into two cells. Each cell had a unique set of 24 jeans. All jeans were labeled by placing cards showing the jean brand, finish and test price in front of each jean.

Respondents were asked to browse through the layout of 24 jeans, just as they would if they were shopping for jeans in a store. They were not allowed to try on the jeans. They were then handed a set of 24 sort cards. These cards were identical to those used for identifying the jeans in the layout. Respondents were asked to separate the cards into piles; one for those jeans they liked and another for those jeans they disliked. Respondents were then asked to rank order the jeans in terms of their purchase interest. The higher the purchase interest, the higher the rank order.

The sample was selected so that the characteristics of those who rated the first set of 24 had similar demographic and jean preference characteristics to those who rated the second set of 24. In this way the ratings and preferences of both groups within market and within age group would be as homogeneous as possible.

Those respondents who sorted the first 24 would be matched with the ratings of the respondents who sorted the second set of 24. Respondents were matched by market, age, and preferred jean brand.

Analytical Method

To estimate the price sensitivity for each jean, the rank-ordered responses from the sort exercise were processed using the Bretton-Clark conjoint utility estimation program.

A cluster analysis of respondent price utilities was performed to measure the impact of price changes within different levels of price-conscious consumers. Since the utilities are estimates of what attributes and features are important to customers, the segments were defined by using the relative importance of each feature to the overall purchase decision. The relative importance for a feature is defined by:

$$RI_{(i)} = 100 \times r_{(i)} / [Sum(r_{(i)})]$$

where;

$r_{(i)}$ = difference between the highest and lowest utilities

Using Bretton-Clark's Simgraf software, market simulations were generated to determine the price sensitivity of the LEVI'S jeans as price varied. These simulations were conducted for each price segment and then summed to measure the total market response to price changes.

Market simulations were estimated for each price segment and then aggregated into the final models.

A variety of choice probability models were used to estimate market share and market share shifts as a result of price changes. Current market share estimates were used as a base to estimate the price elasticity for each product.

Company wide profit simulations were conducted by estimating the impact of each jeans contribution to the company's gross profit margin. By applying the following formula, gross profit margin for each LEVI's garment was computed:

$$GPM_{(i)} = MS_{(i)} \times U \times GM_{(i)} \times P_{(i)},$$

where;

$GPM_{(i)}$ = Gross profit margin for jean (i),
$MS_{(i)}$ = Change in market share for jean (i),
U = Total number of units sold in the market,
$GM_{(i)}$ = Gross margin for jean (i),
$P_{(i)}$ = Change in price for jean (i),

Findings: Consumer Segmentation

Based on the results of the trade-off analysis, consumers were segmented into five groups. They are described below. By conducting a series of cross-tabular analyses of the panel data, the size of each price segment was estimated.

Group 1. High Price Sensitivity: This group of respondents were the most price sensitive and preferred the more basic styles. The jeans they preferred were usually the least expensive.

Group 2. Low/Medium Price Sensitivity: This group of respondents tended to be less price sensitive than the first group and, preferred styles that were slightly more expensive.

Group 3. Low Price Sensitivity: These respondents tended to prefer the widest variety of jeans. More often than the first two segments, they preferred more expensive styles.

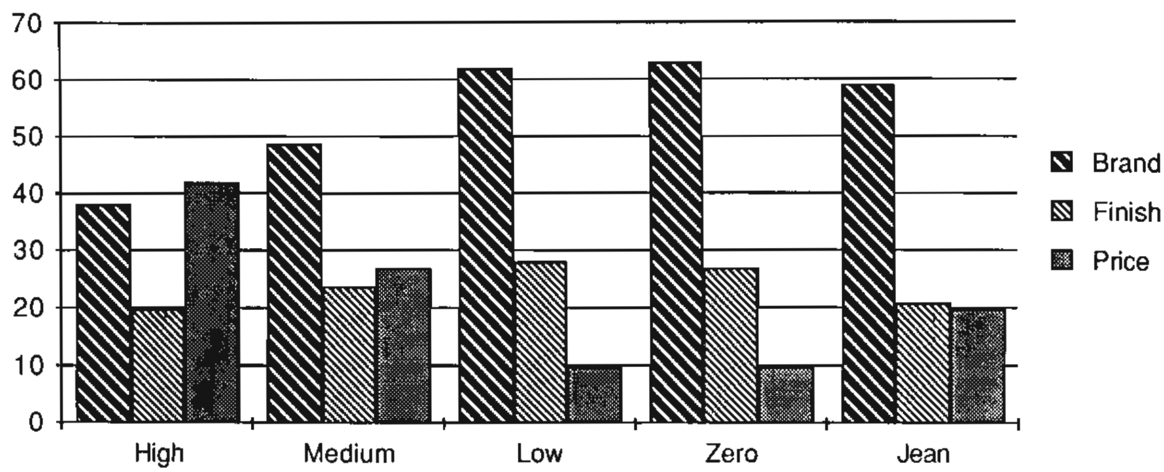
Group 4. Zero Price Sensitivity: These respondents clearly preferred the more expensive styles, and had little concern about price.

Group 5. Jean Driven Zero Price Sensitivity: This group of respondents were driven almost completely by the product. The jeans they preferred were almost always the most expensive styles.

Figure 1 illustrates these results by showing the relative importance of each product attribute in the jeans purchasing decision (for respondents for each segment). Notice how as the importance of price sensitivity decreases, the importance of the jean brand in the decision process increases.

FIGURE 1

Relative Importance Of Product Attributes
In Jeans Purchase Decision



Consumer Price Segments

Market Simulations

A wide variety of market simulations were conducted to estimate the impact of changing prices of different combinations of LEVI's and key competitor products.

Due to the confidentiality of the information being presented, brand names, dollar, and market share changes have been masked. Indices have been created to indicate the magnitude of market share and profit changes due to varying pricing scenarios.

In table 1 below, the prices of Brand 1 garments were raised, while all competitive garments' prices were held constant. The result shows an increase in profit. This illustrates an interesting phenomenon about high-volume industries. In some cases, volume is so great that the increase in revenue offsets the decrease in units.

Notice that there is a certain amount of cannibalization within brand. While certain styles of Brand 1 lose share, other Brand 1 styles' share increase due to the increase in price by other Brand 1 products.

Table 1

Raise all Brand 1 garments price
Hold all competition Constant

Index Values*

Brand	Style	Mkt. Share index	Profit index
1	1	-6	9.4
1	2	0	2.7
1	4	-4.5	8.2

*Indices indicate changes in volume

Another scenario involved the pricing of different brands and styles of Brand 1 garments at a variety of different price levels. Table 2 illustrates what might happen in this situation.

Table 2

Raise Price of Styles A,D in Brand 12
Raise Price of Styles B,C,E in Brand 12
Raise Price of Styles A,D in Brands 1 thru 6
Hold Brands 7 thru 11 Constant
Index Values*

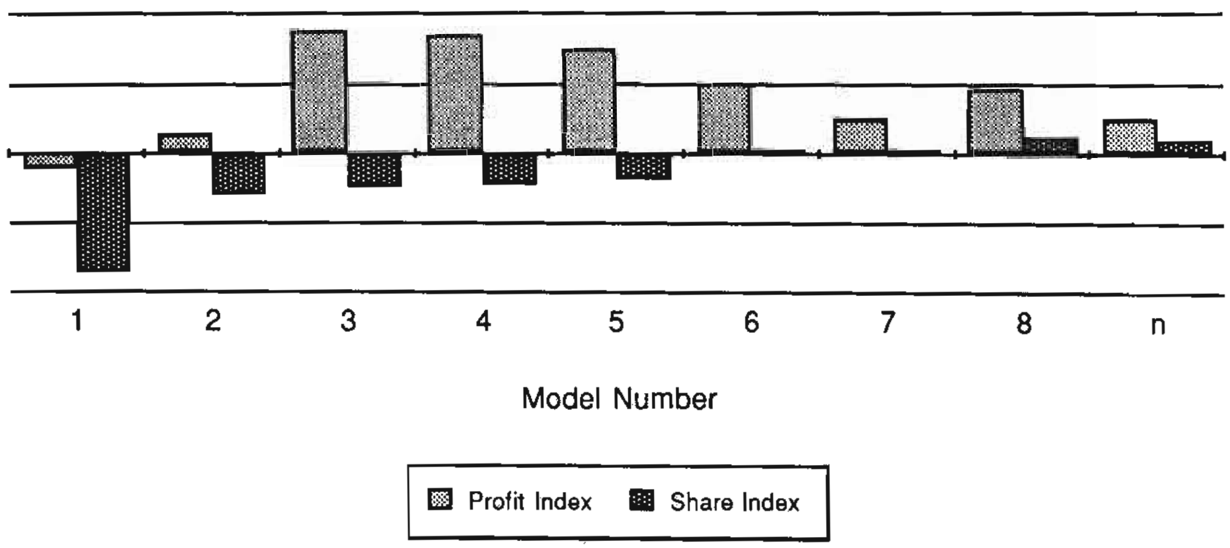
Brand	Style	Mkt. Share Index	Profit Index
1	1	-11	10
1	2	- 7	4.5
1	5	-15	8.4

*Indices indicate changes in volume

Figure 2 further illustrates how a wide variety of pricing scenarios can have varying impact on company profit. As illustrated, in some cases there was a significant loss in both share and profit. In other cases, there was a significant drop in share and increase in revenue; and still in other cases there was an increase in both share and profit. The goal is to choose the pricing strategy that will increase profits the most while minimizing any negative impact on unit sales.

FIGURE 2

Comparison of Profit and Unit Change Indices
By Model



Conclusions

While all these simulations did not lead us to the perfect scenario, they did help Levi Strauss & Co. develop rules to help guide pricing decisions, as well as providing an ability to model potential competitive action with an understanding of their dollar sales and profit impact on Levi Strauss & Co.'s business.

Two important marketing issues were also confirmed by this research:

1. Strategic pricing changes may not be detrimental to a company's profit margin if the decrease in unit sales is offset by a larger increase in revenue due to price increases.

2. Strategic pricing within a product line can cause some cannibalization within brand and increase the unit sales on the lower volume products in the line.

Several additional observations can be made about conducting a strategic pricing study using conjoint analysis.

- Conjoint utilities for each attribute and level are excellent segmentation variables because they illustrate behavioral differences among consumers.

- Using segmentation in conjoint studies also allows the marketer to analyze price and market shifts within different segments.

- The use of readily available statistics such as gross margins, market share estimates, and unit volume will further enhance the research results and allow the marketer to determine the impact of the price changes on the company's bottom line.

Next Steps

Most research is not the "final say" as to what will happen in the marketplace. There are many issues external to the research that can not be controlled for. Conjoint analysis and market simulations are snapshots of a static market and usually can't control for advertising, competitor promotions, and attitudinal shifts in consumer perceptions. Additional testing of any specific pricing strategy should be conducted before national pricing decisions are made. These tests include:

- test markets
- controlled store tests
- laboratory simulations

In addition, the following research studies should be conducted to monitor the impact of any price change.

- Brand image tracking should be conducted before and after price changes to monitor long-term effects.

- Pricing studies should be repeated periodically to track changing price sensitivities due to changes in fashion.

- Retailer surveys should be conducted to estimate retailer response prior to pricing changes.

Comparing Simulated Purchase "Chip" Testing and Trade-Off (Conjoint) Analysis

N. Carroll Mohn, The Coca-Cola Company

In many global markets, marketing researchers have neither budget, software, nor expertise to conduct trade-off (conjoint) studies. Usually the only available option is some form of simulated purchase test. What, then, are the relative strengths and limitations of these two research tools? Further, are results measuring respondent preference the same for both methods? Several lessons can be gleaned from rigorous examination of these issues within the context of soft drinks research.

Both Methods Described

In simplest terms, simulated purchase testing is a controlled experiment. After appropriate screening, eligible respondents are escorted to an interview location with a product display set up. The display as applied to soft drinks includes all major brands of interest. Price labeling cards show current price levels and are displayed in front of each product (i.e., brand/package combination).

Respondents are given ten poker chips to place in front of the various products that would represent their next ten soft drink purchases. Each respondent allocates all ten chips, giving either all of them to one brand/ package or dividing them up among the brand/package alternatives. This serves as the control for the experiment since it simulates the current market situation.

Next, respondents are asked to repeat the chip allocation exercise at a second beverage display in which something new has been added; e.g., a new package for a brand. Since the only change between the control station and test station is the addition of the new package configuration, any share (of poker chip) changes in brands are attributable directly to the new package.

Trade-off (conjoint) analysis similarly begins by asking all respondents some conventional screening questions. If qualified, they also are invited to a central location for the critical portion of the interview in which respondents choice-rank pairs of priced carbonated soft drink products presented to them (again, brand/package combinations). Each respondent reacts to 20 or so pairs of products, uniquely defined by previous responses. Usually pictures are used to illustrate the product alternatives in soft drinks research. The different product combinations are composed of all relevant brands, packages and price points in the market. The purpose of the interview is to derive measures of preference or importance from a subset of the many possible products. The importance measures reflect a respondent's value system of trade-off preferences which, in turn, provide a basis for constructing a market simulation model for estimating market shares under alternative scenarios. (A detailed example of how to construct a trade-off model using soft drinks is contained in "Forecasting Consumer Preferences Using Conjoint Measurement," European Research, July 1982 (N. Carroll Mohn, et al.).)

Relative Strengths

A comparison of the advantages for chip testing and trade-off analysis also gives insight into relative limitations for each method.

Strengths of the chip test cluster along several points.

1. Respondent task is simpler, more realistic.
2. Understanding and communicating the research process is easier.
3. Less research expertise is required for implementation.
4. One study is generally cheaper.
5. Implementation time is shorter.

Alternatively, trade-off analysis has these relative strengths.

1. Required number of total respondents is usually smaller.
2. Many and more complex test scenarios can be explored.
3. Posterior "what if" analyses are possible.
4. Product price elasticities can be quantified.
5. Perceptive insights into market environment interactions are more feasible.
6. Studying multiple issues is less costly.

One comparative point of parity worth mentioning is that each method has an apparatus for establishing statistically reliable results.

Comparing Results

Early in 1988 The Coca-Cola Company conducted three sets of simulated chip and trade-off studies in parallel. Ten studies were conducted in all.

Independent but matched samples of respondents for both research methods were selected randomly, stratified by sex, age, and socioeconomic classification. Screening questions ensured respondents could participate in just one study sample.

The Market I study set included a chip test with 300 respondents for each of three new package configurations being evaluated for Coca-Cola ($n_A + n_B + n_C = 900$) single trade-off study with 600 respondents was used to construct a model for simulating three control and three test stations of the three chip tests.

Tables 1-3 compare estimated market share changes or differences by both research methods. One table is for each of the three new packages tested. The 48 computed share differences generated by the chip and trade-off approaches range from -3.7% to 12.6%. A chi-square test does not allow conclusion of difference between share changes for the two methods at 5% level of significance.

When pairs of differences are subtracted, the resulting difference of differences column in Tables 1-3 is a measure of just how close the chip test and trade-off results are. These difference of differences vary between -2.2% and 1.0%, averaging 0.0%. Signs of the difference of differences do not allow rejection of randomness for a statistical runs test.

Since the new package was added for Coca-Cola only, testing for differences in the share gains from both methods for the brand was a main concern. The hypothesis of equal gains for the two procedures cannot be rejected.

Table	Chip Test Gain	Trade-Off Gain	Calculated Z
1	1.4%	2.3%	0.91
2	6.8	7.7	0.49
3	12.6	11.7	-0.39

Accordingly, results for the two methods are seen as statistically similar for all tests conducted.

Market II and Market III study sets are replicates. In each market a separate chip test with 400 respondents was conducted for two different super liter packages under investigation for Coca-Cola ($n_D + n_E = 800$). Additionally, a trade-off simulation model based on 400 respondents was developed for each test market.

There were two components to the decision situation for these studies. First, management had an a priori commitment to launch either one or the other of the two super liter test packages. This meant that comparisons with the status quo held no interest. Secondly, management wanted direction on how to price the super liter selected. Eight pricing scenarios were specified for investigation with each new super liter package to develop price elasticity curves for both markets.

To make most efficient use of interview time and to accommodate the decision situation requirements, control displays were not used in the chip testing. Instead, multiple test displays were shown to respon-dents, one for each pricing scenario. Moreover, order of presentation of the test displays was randomized.

Trade-off modeling for Market II and Market III followed the same standard procedures outlined previously.

Tables 4-7 compare percentage results from both chip and trade-off methods applied in the two markets evaluating the two super liter packages. One table is for each market/super liter combination tested. The 96 computed differences between percentages generated by chip and trade-off procedures vary from -11% to 12%, averaging 0%. While these differences, perhaps, suggest greater dispersion than those from Market I, a chi-square test at the 5% level again does not justify a conclusion of significant difference. In conjunction, the signs of differences, when subjected to the statistical runs test, do not allow rejection of randomness. When R-squares are computed across studies for each package, the worst case is 94% variability in chip test results explained by trade-off data (and vice-versa). Once more, outcomes from the two research methods are statistically similar.

Implications for Marketing Researchers

Based on the multiple comparative analyses conducted, several useful guidelines emerge. Most importantly, both simulated purchase chip testing and trade-off analysis give approximately the same results, especially where orders of magnitude and market share changes are decision variables in the research.

But when is one method more preferred? Consider use of chip testing when one or more of the following conditions hold.

* Only one or two tests are desired.
* Variable pricing is not a major research issue.
* Decision makers are more familiar with or prefer simpler, more traditional chip testing.
* Trade-off expertise and software are not available.

On the other hand, consider the trade-off approach when:

* Multiple market scenarios are of interest.
* Sensitivity (including posterior) analyses are relevant.
* Price is an important variable.
* Trade-off expertise and computer programs are in hand.

Summarizing by Coca-Cola's experience, both research methods have application. Many opportunities exist for applying either technique -- chip test or trade-off analysis -- to a great variety of substantive problems. Users are urged to pay attention to these guidelines in deciding which method to apply to their specific research issues.

TABLE 1

MARKET I
NEW PACKAGE A FOR COCA-COLA

CHIP TEST VS. TRADE-OFF ANALYSIS
(Simulated Market Share)

Brand	Chip Test Test	Chip Test Control	Chip Test Difference	Trade-off Test Difference	Difference of Differences
Coca-Cola	47.9%	46.5%	1.4%	2.3%	-0.9%
Sprite	5.8	6.3	-0.5	0.0	-0.5
Fanta	13.3	13.1	0.2	-0.5	0.7
Freedom	6.9	7.6	-0.7	0.0	-0.7
7-Up	2.5	2.5	0	-0.2	0.2
Orange Crush	10.8	10.5	0.3	0.2	0.1
BPK	7.4	7.4	0	-0.8	0.8
Ginger Ale	5.4	6.1	-0.7	0.0	-0.7

NOTES:
- o Only the Test percentage for Coca-Cola includes the New Package A, plus status quo packages. The Control percentage for Coca-Cola does not include the New Package A.
- o Chip test sample size is 300.
- o Trade-off analysis sample size is 600.

TABLE 2

MARKET I
NEW PACKAGE B FOR COCA-COLA

CHIP TEST VS. TRADE-OFF ANALYSIS
(Simulated Market Share)

Brand	Chip Test Test	Chip Test Control	Chip Test Difference	Trade-off Test Difference	Difference of Differences
Coca-Cola	53.3%	46.5%	6.8%	7.7%	-0.9%
Sprite	5.7	6.3	-0.6	-1.5	0.9
Fanta	10.8	13.1	-2.3	-1.6	-0.7
Freedom	6.8	7.6	-0.8	-0.6	-0.2
7-Up	2.3	2.5	-0.2	-0.6	0.4
Orange Crush	8.9	10.5	-1.6	-1.9	0.3
BPK	6.5	7.4	-0.9	-0.8	-0.1
Ginger Ale	5.7	6.1	-0.4	-0.8	0.4

NOTES:

- Only the Test percentage for Coca-Cola includes the New Package B, plus status quo packages. The Control percentage for Coca-Cola does not include the New Package B.
- Chip test sample size is 300.
- Trade-off analysis sample size is 600.

TABLE 3

MARKET I
NEW PACKAGE C FOR COCA-COLA

CHIP TEST VS. TRADE-OFF ANALYSIS
(Simulated Market Share)

Brand	Chip Test: Test	Chip Test: Control	Chip Test: Difference	Trade-off Test Difference	Difference of Differences
Coca-Cola	59.1%	46.5%	12.6%	11.7%	0.9%
Sprite	5.0	6.3	-1.3	-2.3	1.0
Fanta	9.4	13.1	-3.7	-1.5	-2.2
Freedom	5.6	7.6	-2.0	-1.6	-0.4
7-Up	2.0	2.5	-0.5	-0.6	0.1
Orange Crush	8.6	10.5	-1.9	-1.8	-0.1
BPK	5.6	7.4	-1.8	-2.1	0.3
Ginger Ale	4.6	6.1	-1.5	-0.7	0.2

NOTES:
- Only the Test percentage for Coca-Cola includes the New Package C, plus status quo packages. The Control percentage for Coca-Cola does not include the New Package C.
- Chip test sample size is 300.
- Trade-off analysis sample size is 600.

TABLE 4

MARKET II
SUPER LITER D FOR COCA-COLA

CHIP TEST VS. TRADE-OFF ANALYSIS
(Simulated Market Share)

Price	Coca-Cola, Liter			Coca-Cola, Super Liter D			Pepsi-Cola, Liter		
Scenario	Chip	Trade-off	Difference	Chip	Trade-off	Difference	Chip	Trade-off	Difference
1	29%	26%	3%	62%	64%	-2%	9%	10%	-1%
2	22	22	0	54	57	-3	24	21	3
3	15	16	-1	81	79	2	4	5	-1
4	11	16	-5	75	72	3	14	12	2
5	51	45	6	45	48	-3	4	7	-3
6	45	42	3	41	46	-5	14	12	2
7	29	31	-2	69	66	3	2	3	-1
8	25	27	-2	66	65	1	9	8	1
			↑			↑			↑

NOTES:

- o Chip test included only Test displays, one for each price scenario for both Coca-Cola and Pepsi-Cola.
- o Horizontal percentages for each method were normalized to add to 100 to facilitate comparisons.
- o Chip test sample size is 400.
- o Trade-off analysis sample size is 400.

TABLE 5

MARKET II
SUPER LITER E FOR COCA-COLA

CHIP TEST VS. TRADE-OFF ANALYSIS
(Simulated Market Share)

Price	Coca-Cola, Liter			Coca-Cola, Super Liter E			Pepsi-Cola, Liter		
Scenario	Chip	Trade-off	Difference	Chip	Trade-off	Difference	Chip	Trade-off	Difference
1	34%	28%	6%	59%	64%	-5%	7%	8%	-1%
2	26	24	2	50	55	-5	24	21	3
3	20	18	2	77	75	2	3	7	-4
4	15	17	-2	73	70	3	12	13	-1
5	55	43	12	42	50	-8	3	7	-4
6	49	41	8	39	48	-9	12	11	1
7	34	31	3	65	65	0	1	4	-3
8	31	29	2	63	63	0	6	8	-2

NOTES:

- Chip test included only Test displays, one for each price scenario for both Coca-Cola and Pepsi-Cola.
- Horizontal percentages for each method were normalized to add to 100 to facilitate comparisons.
- Chip test sample size is 400.
- Trade-off analysis sample size is 400.

TABLE 6

MARKET III
SUPER LITER D FOR COCA-COLA

CHIP TEST VS. TRADE-OFF ANALYSIS
(Simulated Market Share)

Price	Coca-Cola, Liter			Coca-Cola, Super Liter D			Pepsi-Cola, Liter		
Scenario	Chip	Trade-off	Difference	Chip	Trade-off	Difference	Chip	Trade-off	Difference
1	28%	21%	7%	62%	69%	-7%	10%	10%	0%
2	22	18	4	51	62	-11	27	20	7
3	14	11	3	82	82	0	4	7	-3
4	10	12	-2	77	74	3	13	14	-1
5	56	48	8	40	43	-3	4	9	-5
6	49	42	7	32	41	-9	19	17	2
7	30	24	6	68	72	-4	2	4	-2
8	26	23	3	66	69	-3	8	8	0

NOTES:

- Chip test included only Test displays, one for each price scenario for both Coca-Cola and Pepsi-Cola.
- Horizontal percentages for each method were normalized to add to 100 to facilitate comparisons.
- Chip test sample size is 400.
- Trade-off analysis sample size is 400.

TABLE 7

MARKET III
SUPER LITER E FOR COCA-COLA

CHIP TEST VS. TRADE-OFF ANALYSIS
(Simulated Market Share)

Price	Coca-Cola, Liter			Coca-Cola, Super Liter E			Pepsi-Cola, Liter		
Scenario	Chip	Trade-off	Difference	Chip	Trade-off	Difference	Chip	Trade-off	Difference
1	31%	22%	9%	61%	70%	-9%	8%	8%	0%
2	22	21	1	50	60	-10	28	19	9
3	15	15	0	81	79	2	4	6	-2
4	12	14	-2	76	74	2	12	12	0
5	57	46	11	40	47	-7	3	7	-4
6	48	41	7	34	44	-10	18	15	3
7	27	28	-1	71	68	3	2	4	-2
8	24	24	0	70	69	1	6	7	-1

NOTES:

- o Chip test included only Test displays, one for each price scenario for both Coca-Cola and Pepsi-Cola.
- o Horizontal percentages for each method were normalized to add to 100 to facilitate comparisons.
- o Chip test sample size is 400.
- o Trade-off analysis sample size is 400.

Using Ci2 and ACA to Obtain Complex Pricing Information

Greg S. Gum
U S WEST Strategic Marketing

A custom program module permitted each respondent's pricing options to be determined "on-the-fly," based on their answers to Ci2 questions in a disk-by-mail survey. Four specific benefits of the technique were realized: 1) more realistic price points were presented to respondents based on their particular situational context, 2) we had increased analysis capability to better describe and target customers based on their particular feature needs, 3) it allowed us to estimate the impact of potential cross-elasticity with other product lines, and 4) it provided a highly-flexible interview that respondents said they enjoyed.

Introduction

Frequently, researchers seek information on how customers "trade-off" or weigh various product feature options when they consider purchase. These options may include price, vendor/brand, service, or other features of the product. To collect this type of information, researchers may employ conjoint techniques in their studies. Conjoint can provide data for solving many product design/packaging questions, and can provide information for pricing of the product/service. Using these techniques has helped researchers gain powerful information for pricing and designing products based upon input from the marketplace.

Common methods for obtaining pricing information typically include gathering estimates of customers' willingness to pay for the product or service. Ways to obtain these estimates range from directly questioning the respondent in an open-ended fashion to using various scaling techniques (likelihood scales, fixed price points, categorical ranges, unbounded scaling, and so on). Although there are many techniques available for gathering price information, it is frequently difficult to obtain realistic/appropriate price points when researching a new technology product. More often, the respondents may or may not have an initial reference point from which to base estimates of their willingness to pay for a new product.

Current theories on industrial buying behavior emphasize the importance of a reference point when trying to understand the final purchase decision. Recent prospect theory evidence suggests that industrial buyers will evaluate purchase decisions in terms of gains or losses in comparison with some reference point. (Puto, 1987). When researching new technology products, potential purchasers will frequently form their initial reference point based on current pricing of alternatives used or available in the marketplace. So in short, understanding the <u>initial reference point for price</u> was particularly important for pricing of our new product/service.

Furthermore, prospect theory also posits that the <u>perceived context</u> associated with a given decision affects the outcome of the decision making process (Kahnerman and Tversky, 1979). In our study, the reference point varied dependent upon the network context they managed currently. Due to the differing reference points, respondents each had a different initial pricing frame of reference. In this particular study, we tried to address this problem using a special application of Ci2 (Ci2 System for Computer Interviewing) and ACA (ACA System for Adaptive Conjoint Analysis) to provide the respondent with a "customized" initial frame of reference based upon the particular respondent's perceived context (i.e., network design). Our hope was to provide a more realistic and appropriate pricing scenario and therefore obtain better information to do pricing analysis.

Specific Problem

This study involved the development of a new data networking service that required substantial financial capital investment to implement. Because of the potential high risk associated with the decision to enter the market, research information was required. The main objectives of the research were:

1) Estimate the potential demand and revenues for the proposed product/service

2) Obtain pricing and positioning information to better understand the impact of potential competitive entrants

3) Based on customers' product feature package preferences, help prioritize product development efforts

In obtaining the information requested, we encountered a number of unique problems:

- Price estimates for the product could vary substantially, depending upon the respondent's particular situation and therefore required customized pricing

- Poor response rates due to over-surveying

- Cost and interview complexity constraints which did not allow face-to-face interviews.

In this paper, I will focus specifically on the situational-based pricing problem, although some anecdotes of how we addressed the data collection problems will be included.

Methodology

Because each potential customer could have varying situational networking needs, the research methodology had to be capable of identifying and accurately presenting pricing relative to each customer's network situation. By providing the customer with prices relative to his/her network design we hoped to achieve more accurate potential revenue and demand estimates. Further, product management not only wanted to optimize service packaging based on customers' network needs, but also understand the competitive impact of other potential competing solutions.

In solving this unique pricing problem, ACA and Ci2 software and a customized PASCAL subroutine were used to produce a more realistic pricing "scenario" based on the respondent's particular network design. We used the Ci2 module to identify both planned and existing inventories of networking equipment and services. Using the subroutine, each of the respondent's components and usage charges were linked to pricing algorithms.

Estimated costs were then calculated "on-the-fly" based on the most current pricing information for each component/service. The computer then totaled each of the costs to achieve a more accurate overall estimate of the respondents' current and expected network costs. (Previous research studies had shown that respondents typically cannot accurately estimate overall network costs in aggregate.) Finally, each of the cost elements were saved in a database so that they could be accessed for use in the price attribute section in ACA. Further, this information also provided important targeting information for use as segment descriptors utilizing cluster analysis. Based on previous qualitative focus groups and one-on-one interviews with lead users, a revised disk-by-mail survey was designed to obtain the quantitative market information. A pretest was done with a focus group of potential customers. Laptop PCs and the disk-by-mail diskettes were left in the room. Respondents completed the interview with no prompting from the moderator, to simulate the self-administered tasks. The focus group approach was used to:

- Receive feedback about whether the pricing algorithms were realistic and applicable for respondents to base their purchase decision
- Save pretest collection time
- Get actual timing for the length of the new survey
- Allow clients/researchers to view the process

After the respondents completed the survey, an extensive debriefing by the moderator took place to inform the researchers of any specific problems with the instrument or the pricing calculations.

Sample

Because this product required a high level of network sophistication in terms of current equipment and applications present, the eligible universe was limited. Approximately 389 organizations would qualify based on equipment and applications criteria. To simulate response, a post card reminder, phone follow-up calls with an incentive of a rechargeable flashlight, and an incentive of a lottery for a portable CD-Rom player were used. A sample of 126 eligible respondents were "Federal Expressed" a diskette (with express return) with the following results:

78	participated
6	wanted to participate but didn't have access to an IBM PC compatible
9	refused
33	didn't return diskettes in allowed time frame
TOTAL 126	(resulting in a 62% final response rate)

(Other non-computer surveys of similar length with the same population frequently yielded only between 40% - 50% response).

Benefits of Using Ci2 and ACA

Since the incidence of PCs was particularly high with this group of respondents, (i.e., data communications/MIS managers in large businesses and government), we felt that ACA and Ci2 were appropriate given the population studied. We also selected this approach because we had difficulty in obtaining accurate price ranges in qualitative and in-depth interviews. Unless we used some type of "on-the-fly" calculations, we wouldn't be able to pick-up the appropriate prices and present potential customers with a more accurate "situationally-based" price. Additionally, questionnaire length drove us to seek new techniques to increase response from this particular group of customers. And, the questionnaire had to be easy to administer since we were using a disk-by-mail approach to keep interviewing costs down.

Using the "on-the-fly" calculations, respondents could immediately compare their current perceived costs with the new calculated price for the service. Since we were using Ci2, we could also allow the respondents the option to see and change their inventories before the calculations were finalized and pricing was calculated. Prior research had shown that interest and purchase were heavily dependent on pricing relative to current alternatives. Using this module, respondents now had appropriate customized prices based on their current situational alternatives and could complete the trade-off section in ACA.

Aside from the usual analysis capabilities of ACA to simulate potential demand and revenues for the product given various competitive scenarios, another benefit of the approach was the ability to print a "takers" (or high-potential buyer) file. Because many of the sales would have to be customized for a particular customer, the simulator was designed to print inventories, "take" rate or probability to purchase, total estimated costs (current and expected) and other "firmagraphics" that could be drawn from the database for any given simulation. This could then be used to identify high potential "taker" locations and the amount of network concentration within various geographical boundaries. Thus, site selection and product rollout information were provided to give us further insight as to where potential takers may reside.

Since this service was cross elastic with other products sold, we needed to identify specific revenue implications, should we decide to enter the market. By comparing their current costs with the new product/service costs, we gained a better understanding of the amount of revenue won or lost for that particular customer.

Limitations/Next Steps

Although this approach is intuitively appealing, there is little statistical evidence that this approach is superior to other more straightforward techniques. This is not a technique that will be applicable to all products or services. However the author feels that when customization of prices is necessary, more accurate estimates of demand and analysis benefits will be gained from this approach.

A second limitation of the study pertains to the nature of the industrial buying process. Because studies indicate that many industrial purchase decisions involve several individuals (Qualls & Puto, 1989) it is unclear whether we will obtain accurate demand estimates from only one decision maker in the process. Although we were able to provide an applicable frame of reference and contextual situation, we still do not understand the potential effects of the "multiple stakeholder phenomenon" in the final purchase decision.

Other limitations of using simulation techniques are the usual caveats (awareness is assumed to be 100%, distribution and availability of the product is also assumed) associated with conjoint modeling programs in accurately predicting revenues, demand, and share for a product or service.

The author currently is planning on a series of one-on-one validation interviews with high potential takers, non-takers, and non-respondents to validate estimates and better understand the stakeholder process. It is only with continued validation work that researchers can better understand and document the level of predictability of conjoint data used in simulating actual market behavior.

Conclusions

In conclusion, respondents received more realistic prices from which to base their purchase probability estimates, and therefore allowed us to get estimates of price sensitivity and demand.

- The approach obtained market information which could not have been collected as efficiently with traditional mail or phone techniques.
- Analysis capabilities were enhanced and ultimately allowed us to better target and design the service based on customer needs.
- Because of the flexibility and user-friendliness of the interview approach, we received many comments praising the survey technique.

Lastly, the author believes that the technique not only gave more accurate information and estimates, but also can increase and stimulate response with difficult respondents.

References

Kahneman, Daniel and Amos Tversky (1979) "Prospect Theory: An Analysis of Decision Under Risk," Econometrica.

Puto, Christopher P. (1987), "The Framing of Buying Decisions," Journal of Consumer Research, Vol. 3 (December issue).

Qualls, William J. and Christopher P. Puto (1989) "Organizational Climate and Decision Framing: An Integrated Approach to Analyzing Industrial Buying Decisions," Journal of Marketing Research, Vol. 26 (May issue).

HANDWRITTEN DATA ENTRY INTO A COMPUTER

Ralph Sklarew and *Tim Buchholz*
Linus Technologies, Inc.

INTRODUCTION

A new technology has been developed that provides a direct means for handwritten data to be entered immediately into a computer. The data are written directly over the display using a stylus that turns on the display pixels as if the writing were in electronic ink. Because there is a single work surface, single-handed, stand-up, and walk-around applications are possible. Answering questions on an electronic form can be as easy as, or even easier than, on a paper form -- just point at the answer. Questionnaire responses can be recorded without moving a cursor or using a keyboard.

Linus Technologies Incorporated was formed four years ago with a focus on the development and exploitation of the handwriting computer interface. The Corporation's mission is to provide an easy-to-use computer interface that applies a more commonly-used and understood communication practice -- writing.

The first hurdle was the lack of hardware. A computer needed a new input system to track the user's stylus: a digitizer works trough the display or is transparent and fits over a flat computer display. Then, special software needed to be written to allow a computer to recognize what was input. And, finally, a new methodology for data entry needed to be developed because the computer display was more then a paper metaphor, it was automated.

Linus developed the Write-Top as the first computer to utilize this technology. It is a laptop portable, PC compatible that uses a stylus as the input device. It has no keyboard.

HANDWRITTEN DATA ENTRY METHODS

Information and computer commands may be stylus-entered by selection, drawing, writing or gesturing. The easiest and fastest method is by selection of buttons, menu items, "keys," positions or levels, symbols, numbers, or characters. The selection is translated into a character, string of characters, position, command or sequence of commands (macro) that are taken directly by the program or enter by way of the keyboard buffer (pseudo-typing). With stylus (or "pen") on a display there are several methods of indicating the chosen selection: pointing by tapping or touching, or marking by checking, x-ing, or circling. Buttons can be delineated areas of the display that look like a physical button or switch and can show selection by change in terms of shading, color, or text. Buttons are typically used for on/off, yes/no types of selections (though more alternatives may be cycled through).

For more options a menu is used. A menu is a series of buttons where only one can be selected at a time. Menus may be linear or two-dimensional, or successive selections may lead to the display of subordinate menus. The menu may have part or all options visible or may use a button to trigger their display (to conserve display space, simplify the visual presentation, and focus the user's attention).

A key is a button with a simple (usually unchanging single character legend). Likewise, a keyboard is a two-dimensional menu with simple legends. Application-specific keyboards are useful for rapid data entry (e.g., numerical key pads, PC function keys).

In positional or graphical selection, an analog picture of the physical data reading object is used to enter data by pointing or sliding with the stylus. For example, the level of fluid in a thermometer may be indicated on a thermometer display by having a variable shading scheme that follows the stylus. The number of apples and oranges can be graphically illustrated by dragging symbols of each into a displayed box. Digital or analog clocks shown on the display may be set by pointing to reset buttons or dragging the clock's hands. Finally, numbers to be entered into a spreadsheet or database can be drawn on a graph, a back-of-the-envelope data-entry methodology.

Drawing may consist of sketching, sketching with computer-assisted "touch up," or computer-aided design (CAD). For sketching, the display acts like paper with additional features such as editing (e.g., erasing, moving, re-sizing), zooming, and panning. With computer touch up, recognized figures (e.g., circle, rectangle, straight line) are replaced by a computer drawn figure (e.g., a sketched circle is replaced by a computer drawn circle.) CAD introduces another level of computerization: figures can be selected from menus, positioning on a grid can be automatic, and scaling can be general.

Writing is almost universal for data recording. Letters, numbers, and symbols of all types may be written by stylus on the screen as if on paper. In the least constrained writing, termed "scribble," the image or graphic representation is captured by the computer. The image data can be stored, transmitted and re-displayed, or printed. Thus, this method of data collection is ideal for comments which must be kept but need not be manipulated in a data base.

Letters, numbers, and symbols can be written by stylus and recognized by the computer, within limits. The computer is first taught the way the user makes each symbol, including all the desired variations. The requirements for recognition at the present stage of development are that the symbols must be unique, reproducible, and written with a "pen up" between symbols. Uniqueness primarily means that self-similar capital and lower case letters (e.g., Oo, Ss, Pp...) must differ in size, "9" and "q" must differ in position and a single down stroke must be used for only one of "1," "l," or "I." For reproducibility, the requirement is that when an "n" is written that it be closer to one of the pre-taught "n's" than any "h," "r," or other symbol.

Gestures are a special case of symbol recognition in which the symbol is interpreted as a command or macro. Gestures may ink the display as they are written or may be non-inking. They are especially useful for revision or editing where writing a simple symbol may be used to delete, insert, move, or copy a character or a phrase.

LINUS WRITE-TOP HARDWARE

The Linus Write-Top is a laptop portable, PC compatible that weighs 9 pounds and runs for up to 5 hours on batteries. It is constructed in two pieces that are 11" squares. The total thickness is 3.5" with the upper unit a 1" tablet. Only the 3-pound tablet needs to be hand held while in use. The system can be carried in a custom shoulder bag from which the tablet can be extracted for use.

The Write-Top uses a 8088 compatible Central Processing Unit (CPU V20) with Microsoft DOS (Disk Operating System) as the operating system and can run all standard DOS programs. Instead of an on/off switch, it has a sleep/wake-up capability so it can continue without reloading a program and data after an interruption. For internal memory it uses 640 KB of battery-backed CMOS (Complimentary Metal Oxide Semiconductor) static RAM (Random Access Memory), so it does not lose any information when powered down. Additional memory is available with credit card-sized memory cards that can add up to 2 MB for use as a non-volatile RAM disk. Finally, there is a 3.5" floppy disk drive for program and data storage.

The Write-Top uses an LCD (Liquid Crystal) display with optional EL (Electro-luminescent) backlight. The display is CGA- compatible, 640 x 200 pixel resolution in an 8 x 5 inch area. The system has connectors for external power and recharging from AC or 12 VDC auto cigarette lighter, serial communications, printer, external floppy disk drive, phone line (for optional 300/1200 baud modem), and optional keyboard.

The attached stylus is the primary input device. Over the LCD is a panel of glass that is coated for electrical conductivity and that has electronics for detecting stylus position. The glass panel extends beyond the display providing additional input areas. When the stylus is in contact with the glass panel, coordinates of its position are available to the CPU. Typically, these coordinates are used to turn on the display pixels below the tip, providing the look of an electronic ink.

LINUS WRITE-TOP SOFTWARE

Linus software is devoted to four areas: compatibility, services, application generators and applications. Though standard software packages will run on a Write-Top, they are not set up to accept stylus inputs. Keyboard inputs consist of characters, function keys, and cursor-position movements. The compatibility software provides routines to emulate these inputs. For example, when keyboard characters are expected by a program, touchdown of the stylus may result in a lined area appearing on the display into which the characters are written and, when recognized, entered into the keyboard buffer. To the program, this will appear as if the characters were typed into the keyboard. Likewise, if the program was expecting the number of a selected menu item, the compatibility routine that "picks up" the number off the display can be used so that the user merely points to the item number and that number is entered into the keyboard buffer. Function keys can be emulated by designated areas off the display, a pop-up function keyboard on the display, or assigned gestures. Positioning the cursor can be accomplished through emulating the cursor control keys or, if the program is designed to use a mouse, emulating a mouse. Either emulation could be invoked by dragging the displayed cursor or pointing to the desired new location.

The services required to use the stylus are being formulated as a software shell: stylus coordinates, selection, mouse emulation, scribble primitives with editing, character recognition primitives, and compatibility with standard software. These services will be accessible from any application and provide the writing abilities in an easy-to-utilize form. The shell should also help to define a standard "look and feel" for stylus input.

Application generators are programs that make the development of applications as simple as possible. Applications range from simple data collection for which a page image is sufficient to complex programs of error checking, data base management, and logic and/or calculations for which programming is very involved. Application generators span the same spectrum. Simple generators are based on a typed form or interactive screen layout with sequential data output. Moderate complexity is possible with Forms-Write^{tm} which is based on a word processor metaphor and allows calculations, data base management, and communications. For complete generality, Linus also provides Code-Write^{tm}, a set of C-language subroutines for stylus services and display management.

Linus has also developed applications to demonstrate the advantages of the writing interface, such as:

- error checking and rapid data entry for financial traders,
- selection from menus and calculations for ambulance attendants,
- signature capture for package forwarders,
- data input on graphical background for dentists,
- sketching, drafting, and spreadsheet for property appraisers,
- database management and scribble E mail for sales persons, and
- stylus compatibility for applications using standard word processing, database, and spreadsheet programs.

WRITE-TOP FOR MARKET RESEARCH

The Sawtooth Software Demonstration Disk was installed on the Linus Write-Top with the compatibility package. When the demo requires the number of the menu selection, the user points to the chosen number and it is entered into the keyboard buffer. The demo receives the number as if was typed. A template overlay is used off the display to provide carriage return, the end key, and a numerical keypad.

The Write-Top benefits for collection of survey information result from the single work surface, intuitive use, and small package. Combining input and display into the same surface means that the interviewer has fewer places that need attention and greater concentration can be given to the respondent. No "computer literacy" is needed, just pointing is required. Self-administered surveys that even go beyond touch-screen entry with graphical interaction, sketching entries, and scribbled comments are practical. Speed and accuracy are maximized by the immediate feedback inherent in the Write-Top selection procedures. Finally, the small package needs less desktop space and makes stand-up interviews possible.

THE MISSING LINK

PETER A. SCHLEIM
INTERCEPT NETWORK CORP.

Sometime in the immediate future . . .

You are a Marketing Research Manager in a Fortune 500 company. Your brand's major competitor announced a new line extension last week. Your job: obtain consumer reaction to their new product, and determine how yours compares. PRONTO!

You arrive in your office on Monday morning and log on to your computer.

TAP.

Let's see if Winner Research left any messages telling me how they did in getting my survey into the field..

TAP. TAP.

Great! They downloaded the Stimulus and Questionnaires to 6 of the 8 high-penetration markets Friday night and interviewing got started Saturday. Winner also checked their databank for availability in additional cities for similar demographically profiled locations and brought another 2 markets on-line Sunday. Looking good so far. Can't beat that fast a start!

Let me bring up a field status report.

TAP. TAP.

Good! Incidence is on target in all but the California market. The fallout there is on usage question 3 on the screener. Production is behind schedule as a result. Darn!

TAP. TAP.

Oh! Looks like Winner re-balanced quotas over the other cities to adjust for that. And they are running the data to generate an interim "top line" for me this morning. Great! Like they say, "Everybody likes Winner."

TAP. TAP.

I think I'll authorize an electronic debit to Winner's bank account for the full project to show my appreciation for their good work.

Farfetched? Hardly. Except for immediate payment, this evolution is already under way in computer-interactive inter- viewing in malls, and aggressive manufacturers and progressive research suppliers are winning as a result.

Can we identify the environmental forces that are driving computer interviewing towards further evolution? Yes, we believe we can. And are they fundamental changes or temporary? We believe they're fundamental. The forces are outgrowths of changes under way in the three major sectors of the mall intercept industry, and have lives of their own. Things will never be the same again.

• Client pressures.

Client demand for timely/actionable information has never been so intense. And it's very unlikely it will reverse.

DELAYS ACCEPTED --> TIMELY/ACTIONABLE RESULTS ESSENTIAL

- Marginals

- Top Lines

- On-line Data Bases

• Field Agency pressures.

Field Agency restructuring from fragmented, single-site Mom & Pops to regional/national multi-city corporations. is in progress.

FRAGMENTED --> CENTRALIZED

- Availabilities

- Price Quotations

- Bookings

- Project Status Reports

- Financial Controls/Billing

• Research Supplier pressures.

Research Suppliers continually need to differentiate themselves from competitors on several dimensions: capabilities, service, project management and price. Computerized interviewing offers new opportunities.

COLLECTION/REACTION --> PRO-ACTIVE MANAGEMENT

- Completions

- Quota Control

- Terminations

- Incidence

- Length

- CPI/COOP Tracking

- On-line Data Bases

Computerized interviewing in malls has been around for a while, yet pencil-and-paper still dominates. So what changes are going to occur that will add such utility as to cause rapid acceptance and use by marketing researchers?

First, as mentioned above, environmental forces are moving the industry toward a need for "instantaneous" telecommunications. Telecommunications capabilities will render paper and diskettes obsolete in terms of the ability to rapidly gather and process data so that timely, informed decisions are possible.

Second, digital video and audio will displace hard copy stimulus enabling immediate, inexpensive duplication plus consistent, accurate presentation.

What will drive these changes? Need and Greed.

• Economic pressures

Computer-assisted interviewing systems can yield cost reduction opportunities (though not necessarily result in net savings). The most efficient computer interviewing solution will be one that provides the opportunity to heavily impact the major cost centers without sacrificing quality.

TYPICAL PROJECT COST CENTERS

	% of Total
Field	57
Stimulus	12
Project Control	11
Shipping	8
Printing	6
Keypunch	5
Telephone	1
Total:	**100**

Ordering the cost centers of a project provides a clear picture of where the meaningful cost-savings opportunities exist. Ideally, the new generation of systems should reduce the following costs:

1. Field Interviewing

Any "ideal" system will enable self-administered interviewing through a user-friendly interface and clear, "on-system" instructions. The system should optimize respondent interactivity, the goal can be total interviewer displacement.

2. Stimulus

Any "ideal" system will be able to present and control the stimulus as well as the questionnaire process. Other goals are to totally displace costly and time-consuming stimulus duplication efforts.

3. Project Management

Any "ideal" system will automatically provide daily and therefore actionable - field management reports. The goal is to displace the wasteful effort of collecting management information and instead use those resources to analyze and act on the information.

4. Shipping/Printing/Keypunch

Any "ideal" system will support electronic distribution and retrieval of field materials/ completed questionnaires, and update and manage a series of remote interviewing sites. The goal will be to displace the exchange of all hard copy materials by more efficient and faster telecommunications.

The conclusion is almost inescapable. To capitalize on timing and cost opportunities, the next evolution of computer interviewing <u>must</u> also ...

enable remote telecommunications

and

integrate an automated ability to present stimulus

Well then, why don't we all go do it?

Both your PC and your telephone deal in the same stuff an electrical signal over a wire - so you might expect that connecting them would be a simple matter: you need a modem, you need communication software, and you need someone to call.

Sounds simple, right?

Wrong!

Instead, you find you've fallen down an Alice-in-Wonderland rabbit hole and entered a world where you have to choose from hundreds of modems; all operating under different speeds, different protocols, and many incapable of even communicating with each other.

Let's take a very simple file transfer and see what decisions need to be made on both sides of the transfer:

1. What speed modem am I calling into?

    300 bps
    1200 bps
    2400 bps
    9600 bps
    19200 bps

2. What Parity are you?

    ODD
    EVEN
    SPACE

3. What is the Stop Bit setting?

    1 Bit
    2 Bits

4. What is the Data Bit setting?

    7 Bits
    8 Bits

5. What transfer file protocol will we be using?

    ASCII
    XMODEM
    WXMODEM
    TELINK
    YMODEM
    MODEM7
    KERMIT

6. What phone number will I be dialing into?

7. What security pass word am I authorized to use?

8. What is the name of the file that I will be picking up and in what directory is it?

9. Are you in communications now or do you want me to dial in at a specific time when your computer is free?

Now multiply this by all the markets you use. It's the Tower of Babel come to pass:

"And the Lord said, 'Behold, they are one people, and they all have one language....Come, let us go down, and there confuse their language, that they may not understand one another's speech.'" Genesis 11.1-9

So much for simplicity.

Then there are system and operational issues that need to be addressed:

1. Modem Protocols, Speed
2. Communications Software
3. Attended/Unattended
4. Local Area Networks for Multiple Station Cities
5. Dedicated Lines vs. Voice Grade Lines
6. On-line vs. Batch Transmission
7. Security Access
8. Maintaining Dialing Directories

Having made it this far you are ready to start. Now comes the realization that you are into a 24-hour operation because to be price competitive, you will want to take advantage of after-11 night rates for transmissions. And you will need to process data overnight to provide service to your clients who will increasingly expect to have the prior day's interviews available on their computer the following day. Mounting competitive pressures necessitate this sort of turnaround and information availability.

Now that you have set up the communications capability let's turn to digitized stimulus. If you thought that you had just visited the building site of the Tower of Babel when exploring Telecommunications, you are about to find out that hooking up an analog video with the digital computer is like striving to obtain consensus amongst the Tower of Babel's engineers. The only standard that you can be sure of in this area is that no standard exists.

First, the process of capture and display is hardware intensive. And converting an original image to a digitized image requires a highly-skilled video technician. Once again you are faced with a daunting array of choices for:

Input Devices

Capture Hardware

Display Monitors

Color Resolution Standards

Data Compression Software

Image/Audio Management Software

Disk Storage Issues

Time and space prohibit an adequate review of this subject. And, it is highly unlikely this has anything to do with the business of the readers. This is simply not the business of Marketing Researchers.

Aside from the discussed complexities, fundamental barriers exist within the industry, thwarting these needed changes. These barriers emanate naturally from the competitive nature of the industry. Cooperation and industry-wide participation are required to justify the cost of such an undertaking. It is simply unnatural for any manufacturer, fullservice research company, or field agency to provide these services to the rest of the industry. It's like expecting a fox to live in a hen house without snacking on a drumstick from time to time.

Our belief is that independent, third-party companies need to be the service providers offering an open system for all suppliers and all field agencies in support of all client studies. These service providers will set standards, deploy compatible hardware to the field, manage the programming/digitizing/telecommunications operations centers, generate project management reports, give technical support to all users, and provide training to interviewing staff.

While there are certainly barriers to evolving to the next stage, they are already starting to fall away. Competitive manufacturers, research suppliers, and field agencies are already positioning themselves to be tomorrow's winners. They're gaining experience today in computerized interviewing and are now ready for the next generation. There is plenty of room for winners in the new world; those who adopt and adapt to the changing environment. As with all evolution, we are seeing the survival of the fittest in the world of marketing research. We believe that the intelligent use of technologies is the best path to that survival.

LARGE-SCALE CONJOINT DATA COLLECTION AT THE CHICAGO AUTO SHOW

Linda S. Middleton
Chicago Tribune

When I am asked what I like about the marketing research environment at the *Chicago Tribune*, I respond that I like our approach to research projects: the creativity and the willingness to try something new.

This year the *Chicago Tribune* tried something new with their annual survey at the Chicago Auto Show. Actually, we tried several new things.

First, we took our traditional paper and pencil survey and computerized it. This, in itself, really caught people's attention.

But then we also tried a new interviewing technique called, "Adaptive Conjoint Analysis." Now, to the research community this might not seem like anything new or different, but to the end users, a group of advertising sales executives, this was different.

The purpose of this paper is to discuss why and how this project was executed and how Adaptive Conjoint Analysis is working for the *Chicago Tribune*. This paper will also show the relative ease with which a similar project could be completed and thus produce quality data at a reasonable price.

Specifically, this paper will address three questions: Why did the *Chicago Tribune* conduct this study? How did we set up and execute the study? What was the outcome?

WHY DID THE *CHICAGO TRIBUNE* CONDUCT THIS STUDY?

The *Chicago Tribune* generally completes Marketing Research studies to accomplish one of two objectives. The studies either provide strategic direction for the *Chicago Tribune* itself, or they generate information to support our advertising sales efforts. For this project, the latter applied.

The research client for this project was the Automotive Advertising Sales Staff which services automotive dealers, dealer associations, and national manufacturers. Their need was for new and detailed Chicago market automotive data.

As with most *Chicago Tribune* "advertising" research projects, the automotive staff wanted this study to provide:

* information that is of value to the advertising client,
* information that differentiates the *Chicago Tribune* from other media companies,
* and information that supports a media recommendation.

The staff wanted to interview those who had recently purchased a new vehicle, as well as those who were planning to purchase. They had prepared a long, detailed list of requested information and had a budget of approximately $10,000. We began by evaluating alternative courses of action.

Our first option was to do nothing, which, in this case, was not a viable choice.

The second option was a random sample survey of New Vehicle Purchasers. The sample would be drawn from new vehicle registrations within the Chicago area. The interview would be conducted either over the telephone or through the mail. One advantage of this technique was that the random sample would provide representative and projectable market information. However, this solution also had several drawbacks. First, and foremost, with the required sample size and the volume of data requested, the cost of this study would exceed our budget. Second, we would be limited to interviewing only those who had recently purchased a new vehicle, not those who were planning to purchase.

Third, neither a mailed or telephone survey would meet all our needs. A mailed survey would allow us to obtain the detailed information we required, but would limit our use of skip patterns. The telephone survey would allow more intricate skip patterns, but would limit the amount of information that could be obtained.

Another proposed option was to continue gathering data as we had in the past: A paper and pencil survey at the Chicago Auto Show.Since 1984, the *Chicago Tribune* had interviewed attendees at the Auto Show using a pencil-and-paper method. Approximately 2000 adults completed a one page questionnaire. While the information was limited, automotive manufactures and dealers liked the survey and the information it provided.

The *Chicago Tribune*'s Automotive sales staff liked the idea of interviewing at the Auto Show. However, they were not in favor of repeating the pencil-and-paper survey because it would once again limit the amount of information gathered.

Our final option was a computerized survey at the Chicago Auto Show. The benefits of this option included both the ability to obtain detailed information and to use complicated skip patterns. The use of computers was also favored because it would be new and creative.

In evaluating our options, I felt that we should continue to take advantage of having a large group of our target respondents in one central location. The Chicago Auto Show is the largest automotive show in the United States, drawing close to one million attendees each year. The 1989 show ran for nine days, from February 11 through 19, and was open from 11:00 a.m. to 11:00 p.m. daily.

While the automotive staff had been opposed to another pencil-and-paper survey at the auto show, they liked the idea of a computerized one. Our decision was made: we would execute a computerized survey at the Chicago Auto Show.

As the questionnaire development progressed, two factors impacted our direction. We decided that we wanted to know what product attributes were important to prospective buyers in choosing a particular vehicle. We also became intrigued by Sawtooth Software's Adaptive Conjoint Analysis (ACA) package.

Therefore, a second decision was made. We would conduct two different surveys at The 1989 Chicago Auto Show: The New Car Buyer Survey and The Prospective Car Buyer Survey.

The New Car Buyer Survey included only Ci2 (Ci2 System for Computer-Interactive Interviewing)-programmed questions and took the respondent approximately fifteen minutes to complete. This interview focused on the behavior of the new car buyer. The questions encompassed the process of selecting and acquiring a new vehicle.

The Prospective Car Buyer Survey focused on consumer preferences for product attributes. It was, with its use of Adaptive Conjoint Analysis, by general consensus, "the fun survey."

On average, this interview took approximately twenty minutes to complete. The questionnaire included both Ci2-programmed questions, for segmenting purposes, and ACA questions. The ACA portion of the questionnaire measured 18 attributes related to the purchase of a vehicle. External visual aids were not used;the entire survey was self-contained within the computer.

HOW DID WE SET-UP AND EXECUTE THE STUDY?

From the early design stages of this project, we worked closely with the organization that sponsors the show, the Chicago Automobile Trade Association (CATA). We agreed that the survey booth would not be identified as a *Chicago Tribune* function, but rather as a part of the show itself. A banner simply identified the booth as the "1989 Automotive Survey."

There were several benefits to this decision. The respondents perceived the survey to be a part of the Auto Show and one that would benefit all dealers and manufacturers. Also, by masking the *Chicago Tribune*'s involvement, we would not bias their responses. It also allowed the CATA to turn down numerous requests, including those from our competition, for entrance/exit surveys. Our survey was considered the "official" survey of the auto show.

The Chicago Automobile Trade Association provided the exhibition space for the survey booth, which measured 24' x 14' and was located in the lobby area of the main entrance to the Auto Show. It proved to be an outstanding location because the visibility and traffic flow provided a constant pool of respondents.

The survey booth was built using skirted tables as barriers and had blue carpeting within the interior. Ten IBM PC's (640K, dual floppy, color monitor) were placed on the tables and used as the interviewing devices.

As an incentive to complete the survey, each respondent was entered in a drawing and eligible to win one of several prizes. The entry blank for the drawing was the last question of the survey and respondents were informed that the computer, at the conclusion of the Auto Show, would randomly select the winning entries. Two signs were posted within the booth that identified the prizes that the respondent was eligible to win.

To protect against any theft of the computer equipment, a security company was contracted to guard the survey booth during the hours that the show was not open.

Hiring personnel to staff the booth turned out to be our biggest challenge. Time, ability, and monetary compensation were the basic concerns in hiring a staff. We needed interviewers who were capable of managing the crowds, were willing to work for nine days straight, and whose salary would fall within our budget. Interviewers also needed to be dependable and to present a positive image.

After searching internally and approaching several outside research firms, we hired our interviewers through a temporary employment agency. All of the interviewers attended a training session prior to actually working at the Auto Show survey booth.

We scheduled two shifts of workers daily. Each shift included a supervisor - or "point person" - and several assistants. On Monday through Friday, we employed three interviewers per shift. On the weekend days, when the crowds were larger, we used four interviewers per shift.

Each worker performed one role. They either acted as the "screener," the "gatekeeper," or they provided instructions to the respondents.

The "screener," using a standardized procedure, stood outside the survey booth and qualified potential respondents for the appropriate survey. Respondents were screened to determine if they were either recent or prospective car buyers. The New Car Buyer Survey required that the respondent had purchased or leased a new vehicle within the past 12 months. The Prospective Car Buyer Survey required an interest in purchasing or leasing a new vehicle within the next 12 months.

We also screened out respondents with particular knowledge of or interest in the automotive industry. Respondents were disqualified if they worked within the automotive industry, or for a media company, advertising agency, or marketing research firm.

With regard to the screening process, it is notable that we did not have to solicit respondents. Generally, respondents waited in line to take the survey.

The "gatekeeper" monitored the entrance to the booth and allowed only the actual respondents within the booth. To maintain order within the survey area, children, spouses, and friends had to wait outside the booth.

Once a respondent was screened and a computer was available, one worker would then be responsible for seating the respondent and providing necessary instructions.

This person also maintained the computer operations, as necessary. Basically, this involved changing interview disks as they were completed. The disk was programmed to allow 10 completes per disk, to generate automatic respondent numbers, and to use a respondent number message screen.

Although there was a concern that a computer breakdown or disk failure might occur, the computer operations ran smoothly. We did not incur any down time because of computer failure, and we did not lose any completed interviews due to disk failure.

WHAT WAS THE OUTCOME?

The survey booth processed close to 2600 respondents during the Auto Show. A total of 1572 adults completed the Prospective Car Buyer Survey; 60% were male, 40% were female.

There was a slight skew to younger age groups, but this seemed representative of adults who attend the Auto Show.

Age groups:	18 to 24 - 19%
	25 to 34 - 37%
	35 to 44 - 24%
	45 plus - 20%

Respondent reaction was extremely positive to both the survey and to our use of the computerized-interviewing technique.

The respondents showed high levels of curiosity and willingness to participate. In fact, many said they felt that because the interview provided insight into their personal preferences, they actually benefitted from participating in the survey.

Likewise, the interview, with its colorful screens and computer interaction provided enjoyment. This enjoyment translated into a strong level of commitment from the respondent. Less than one percent of the respondents terminated their interviews before completion.

The nature of the interviewing process also provided a high degree of involvement. Many observers of the survey booth commented on the levels of intensity and concentration that respondents displayed. The majority had little, if any, problem understanding the task that was presented or had any trouble completing the survey. In fact, analysis of the results showed that 85% of the completed surveys had a correlation factor of.800 or higher.

Overall, the extremely positive reaction by respondents provided the *Chicago Tribune* a large quantity of high quality data. By contrasting the quality and depth of the data obtained to our expenditures, this study proved to be a good value for the money spent. The actual cost of this project was approximately $10,050. An itemized look at our expenditures follows:

Computers	$ 1,800.
The booth	1,800.
Interviewers	4,200.
Security	1,200.
Field disks	200.
Incentives	850.
	$ 10,050.

HOW IS THE CONJOINT INFORMATION BEING USED?

The *Chicago Tribune* has begun the process of sharing the conjoint analysis data collected at the Auto Show with our advertising clients and, to date, the response has been very positive.

Because of the significant sample size, we can customize the analysis and presentation of Auto Show results to a particular account. We can analyze specific market segments and how the preferences change depending on the target. For example, using a minimum correlation factor of .800, the following details what the available sample would be for each respective target:

Target	Sample
Adults, aged 18 to 34	780
Men, aged 35 to 44	192
Single, college-educated adults	132
Married adults, aged 25 to 44 with no children	159

Working directly with dealers, dealer associations, or automotive manufacturers, we can customize the conjoint analysis to meet their needs. This ability also continues the building of partnership relationships with our advertising clients.

Furthermore, the simulation capabilities of the Adaptive Conjoint Analysis program allow us to play "What if ..." and to present challenges to our advertising clients:

Give us a market and we will find you a product. For example, what combination of attributes is preferred by adults, aged 45 to 54, with a household income of $35,000 or more?

Give us a product and we will find you a market. For example, what age and income of adults would be a prime target for a mid-size car priced $20,000, with a T-roof, automatic transmission, and cloth seats?

Is your advertising pushing the right buttons? For example, are you including, within your newspaper advertisements, product attributes that are important to your target audience?

WHAT WOULD I DO DIFFERENTLY NEXT YEAR?

Because this was the first time that the *Chicago Tribune* had attempted such a research project, horror stories about the actual execution might have been expected. But the reality was that this project ran smoothly with very few problem areas. However, as with all research studies, there were a few areas that we perhaps could improve upon. These areas included aspects of the questionnaire, the screening, and the staffing.

Questionnaire

I would include external visual aids for the respondents. For example, I would use actual color charts to measure the preferences for color combinations. This would aid in the analysis of the data because we could then address preferences for an actual shade of a color rather than for a generic one.

I would also work on the visual presentation of several of the ACA questions. Most notably, people became confused during the preference ranking section. They would answer their first choice, but not realize that they needed to continue. I would like to change the colors on the screen or use some kind of visual flag to aid the respondent.

I would limit the timing of the survey to 15 minutes. Even though the length of the questionnaire did not discourage respondents from completing the survey, this would allow us to turn the computers over at a faster rate and reduce the frustration of those waiting in line.

I would also, if possible, increase the number of interviewing devices. As we could have easily kept fifteen to twenty computers occupied, the incremental cost of additional computers would be minor compared to the increased number of completed surveys.

Screening the respondents

Even though our screening process worked well, I would like to improve our procedures. Because the interest in completing the survey was so great, I would consider attempting a randomization process to select respondents.

Staffing

I would hire workers for a particular role or job function rather than as a "generic staff." For example, in hiring a worker to perform as the screener, I would look for someone who presented a positive image, was able to answer diverse questions, and to maintain control over the screening process. The gatekeeper, on the other hand, is a less demanding position.

I would also hire more workers, so that each worker could have one day off during the show. This would reduce the fatigue factor associated with working nine days straight.

CONCLUSION: THE IMPACT OF THIS PROJECT

The 1989 *Chicago Tribune*'s Automotive Survey was successfully executed and has been positively received by the automotive sales staff and their clients. This project successfully accomplished the objectives of providing:

* information that is of value to the advertising client,
* information that differentiates the *Chicago Tribune* from other media companies,
* and information that supports a media recommendation.

By completing in-depth interviews of 1,600 prospective car buyers and 1,000 new car buyers, the *Chicago Tribune* developed an extensive database of Chicago Area vehicle purchasers. This data will allow the *Chicago Tribune* to provide information that meets the needs of their advertising clients.

The quality and nature of the conjoint data now available to advertisers will continue to enhance the *Chicago Tribune*'s image as an innovator of marketing information.

Likewise, the quantity and uniqueness of the information available will continue to increase significantly the *Chicago Tribune's* ability to make effective sales presentations and media recommendations.

COMPUTER INTERVIEWING APPLICATIONS IN THE NAVY

Emanuel P. Somer and *Dianne J. Murphy*
Navy Personnel Research & Development Center

For the Navy to meet its manpower objectives and goals for the 1990's, new and creative people-oriented programs and policies are being initiated, maintained, and evaluated. To achieve this, more attention is being placed on the needs, attitudes and opinions of Navy personnel. The primary method of obtaining this information is through the use of personnel survey technology. To meet the challenge the Navy has endorsed the implementation of the Navy Personnel Survey System (NPSS).

Under NPSS, work is continuing to improve the quality and timeliness of personnel survey information, to reduce intrusion on Navy commands, reduce possible redundancy, and reduce data collection costs. This is being accomplished primarily by the development of sophisticated sampling strategies and the increased use of computerized interviewing surveys.

Within the historical context of survey research in the Navy, the use of computer interviewing technology is relatively new. We began as recently as 1985 by using computer technology to measure attitudes and opinions. In 1985 we were tasked with determining if we could use a computerized survey network to conduct quick reaction surveys, primarily on Navy civilian personnel. The project was called CENSUS (Computerized Executive Networking Survey System).

When CENSUS began in 1985 it was conceived as a networking system. The CENSUS System I prototype was configured as a multi-user survey system with up to eight individuals working at their own speeds connected to an IBM PC/XT or PC/AT host computer with 640K RAM. The system software partitioned 30K of the host's memory to each of eight terminals, which were linked with the host directly, or remotely through modems. Northern Telecom Displayphones served as terminals for CENSUS I. Specially designed multi-user communications software was used and respondents' answer files were stored in ASCII format on the XT or AT. While it demonstrated the potential capabilities of such a system, a number of hardware limitations forced the evolution to CENSUS II.

CENSUS System II was configured with an enhanced IBM/AT containing 3Mb of memory and 16 serial ports. Connected to these ports were sixteen Qume ZVT211GX terminals, allowing a greater number of respondents to take a survey during any given administration period. The CENSUS software was programmed at the Navy Personnel Research and Development Center (NPRDC) in the C programming language. The advantages of the system outweighed the drawbacks of the previous prototype. The system proved to be especially useful in situations where the CENSUS host computer and terminals were "hard-wired" in one location, and when large numbers of individuals were required to complete an automated survey quickly.

CENSUS System III was the first stand-alone version of the CENSUS system. It utilized the standard floppy diskettes to run the survey on any IBM PC or PC-compatible computer. The software called MASQ (Microcomputer-based Assessment, Surveys, and Questionnaires) was developed at NPRDC using the COBOL programming language. System III evolved in an effort to eliminate the need for a computer network to administer CENSUS surveys. Once again, limitations in the software and advances in available hardware and software allowed the evolution to the current state-of-the-art, CENSUS IV.

CENSUS System IV is the existing prototype for NPSS computerized surveys. It utilizes the Sawtooth Ci2 System (Ci2 System for Computer-Interactive Interviewing) to conduct the surveys, often using diskettes as an alternative to pencil-and-paper in mail surveys.

The newest expansion of System IV is currently in the testing phase. It includes the development of a six-line bulletin board system using TBBS (The Bread Board System) software. The 24-hour BBS will allow assigned individuals at remote sites to download surveys onto floppy diskettes. These individuals will then be responsible for distributing the diskettes to a pre-determined sample, coordinating the survey administration, and uploading the data file to the Center through the BBS.

The initial test of the BBS is being conducted using Navy civilian personnel. During the summer of 1989 two short "shakedown" surveys are being conducted to work out any problems of implementation, particularly those encountered with the "human element" of the system. It is anticipated that the most difficult problem to overcome in the establishment of such a sophisticated delivery system is the development of a reliable coordinator network.

After initial de-bugging of the system, effort will be focused on the administration of the annual trend survey of Navy civilians using the BBS. As mentioned previously, the survey is a cross-sectional, omnibus survey addressing aspects of job satisfaction, as well as other issues of relevance at the time of administration. It is to be administered yearly to a sample of approximately 2,500-3,000 Navy civilians, including blue-collar workers, at 128 sites around the world.

Clearly, the logistical problems to be overcome in developing such a system are immense. Experience thus far has shown that there are a wide range of computer expertise and motivational levels within the coordinator network. Some people have virtually no experience with a desktop computer and modem. Further complicating the delivery system is the large number of civilian employees, including shipyard workers and construction employees, who have neither ready access nor previous exposure to a personal computer.

Despite the problems, it is anticipated that the advantages to such a system will far outweigh the disadvantages. Obviously, the greatest advantage, particularly when dealing with government services, is the elimination of postal delays. It is estimated that this alone will result in a time savings of about one month during the administration period. Further time savings occur when considering that the keypunching and data cleaning functions that are eliminated with the use of Ci2.

Experience thus far within the CENSUS framework using computer interviewing technology has shown the response to be overwhelmingly positive. The Navy as a whole and the research and development community in particular, strive to be very "high-tech." The difficult portion of the task is, and will continue to be, the "human element" - getting a reliable coordinator network and getting the job done with a minimum of intrusion into daily work routines.

Now, we'd like to share with you another Navy application of computerized surveys - administration of a survey at a Navy-sponsored conference. As we take you through the process, we'll share with you the things that worked well and those that didn't work quite so well. It is hoped that in so doing you can learn from our experiences.

In November of 1988, the Navy conducted a Family Support Services Conference in Norfolk, Virginia. This was the second conference of this nature ever conducted by the Navy, the last one being 10 years before. It was attended by approximately 1,300 high- ranking Navy and Marine Corps officials, including about 70 of "flag rank" or equivalent. The objectives of the conference were to align family support efforts with future needs of operational commands and to establish family support goals and strategies for the year 2000.

Our task was to collect data during the conference and feedback the results to attendees in a formal briefing during the closing session. The content of the survey instrument included questions directed toward predictions about the future in general and the impact of potential changes on the Navy in the year 2000.

We were tasked with gathering data on-site using the computer for a 50% sample of attendees, while the other 50% completed an equivalent paper-and-pencil version of the questionnaire. This was our first mistake: Do not offer a pencil-and-paper alternative to the computer questionnaire, at least until after the end of the computer administration. For those who are computer-shy, an alternative is a way out!

Determining the content of the survey instrument was a team effort and was accomplished in a minimum of time. A much larger block of time was spent "computerizing" it and making it "user friendly." Every screen was analyzed and re-analyzed to determine if the color combinations worked, if the right words or statements were emphasized, and if the ambiguity in response mode was eliminated. We tried to take nothing for granted. We can't stress this enough: It is extremely important to make the survey user friendly, particularly when the respondents are unfamiliar with computers and/or are high ranking officials who could potentially feel embarrassed by their lack of computer knowledge/expertise. In creating the computerized version, we also tried incorporating humor in a few of the screens to allow some "comic relief" from the intensity of the survey. Most respondents appreciated our attempts, even if the humor was less than side-splitting.

The logistics of the effort were, at best, challenging. We were, admittedly, nervous. (After all, this was our first use of the Sawtooth interviewing software, and we were doing it on such a grand scale. Expectations were high and so was the visibility of our success - or failure!) With the conference leaders anticipating approximately 500 survey subjects, we determined that we would need 40 personal computers to provide adequate coverage. However, the conference was on the East Coast and we were on the West Coast. How could we get our hands on that many PC or PC-compatible computers?

We located a computer rental agency on the East Coast which could help us with our problem. From them we rented all of the administration, analysis, and report preparation equipment that we needed to complete the task. They provided delivery, set-up, technical support, and teardown after the end of the conference. We simply arrived with our software (and everything else that wasn't tied down) and went to work. Renting equipment from a large-scale professional organization worked very well, and we would highly recommend it. It greatly reduced the workload required in setup and allowed the experts to deal with the logistics and all the possible hardware problems.

Because of conference size, the accommodations encompassed three hotels. We decided to place the larger number of computers (24) at the main hotel and create satellite computer sites at the other hotels. This leads to our next big mistake: Do not depend on the hotel to know their electrical system. Double- and triple-check power sources for potential overload when all computers are on. We lost a few respondents on that one!

Each of our survey centers was manned at all times. This seemed important to the respondents, - even though the survey guided them through the process, having a friendly face to get them started helped ease anxiety. We also suggest that survey center hours be established as soon as possible and that they be clearly posted. Once established, do not change them. Run the survey centers as a merchant would run a store. Likewise, good communications capabilities, either by telephone or walkie-talkie, should be established among the multiple sites.

The survey center in the main hotel created a "high-tech" environment, lending credibility to the administration of the survey by the "demand characteristics" of the environment. Although it is doubtful that these demand characteristics affected in any real way the number or quality of the responses for this particular survey, it very likely had its effect by promoting the Center as a high-tech, state-of-the-art survey center.

We decided by mid-conference that the location of the survey center in the main hotel was not optimum - respondents had to go out of their way to find us. This brings us to another point: Whenever possible, be actively involved in the planning of the entire conference, including logistics. This will provide an opportunity to develop a "presence" at the conference. This presence can be established by (1) locating the survey administration center in a high traffic area, such as the lobby of the hotel or equivalent, (2) allowing scheduled times in the program that are dedicated to survey administration only, and (3) allowing increased access to media/publicity opportunities.

Because participation in the computerized version of the survey was voluntary, we tried a number of ideas to increase interest. One idea that worked well was a specially designed and created "I Took It" sticker that was given to each respondent who took the computerized survey.

We closed down the survey centers at six o'clock on the last full day of the conference. We then began the all-night analysis session, ending in the presentation of a formal briefing to all conference attendees. This was a pressure-packed situation, and one that taught us many things about future conference applications. Next time we'll: (1) rent a suite, hotel-style apartment, or separate hotel room to serve as the home base/analysis station; (2) begin the analysis with the first wave of data so that there is time to "process" the data and develop a story line without being in a time crunch; (3) if possible, allow more time for final analysis, perhaps closing the computer centers down 24 to 36 hours before the outbriefing; (4) use the preliminary data base as a straw proposal for the outbriefing and meet early with conference officials to determine the tone/direction of the briefing materials; and (5) conduct a team meeting/critique of procedures, methodology, etc., as soon as the doors are closed.

All in all, the computerized survey was very well received by the conference attendees. As we look toward the future and other applications of computerized survey technology, we continue to learn more. We hope to expand our capabilities to applications even more sophisticated techniques.

NOTES

The opinions expressed in this paper are those of the authors, are not official, and do not necessarily reflect the views of the Navy Department.

This manuscript is submitted with the understanding that the right of reproduction for government purposes is reserved for the Navy Personnel Research and Development Center.

Computer Interviewing in Europe: What U.S. Researchers Need to Know

Martin Stefflre,
PA, London, England

Computer Interviewing works abroad! All the advantages of using PC-based consumer interviewing still apply. Unhappily, the disadvantages are there too. The disadvantages are somewhat magnified because of the lack of field interviewing services trained in using PCs and because PCs generally aren't as widely used as they are in the US.

PA Consulting Group is a multi-national consulting organization. The part that uses computer interviewing is the Strategic Services Group of the Marketing Division. Their clients range from small specialized manufacturers of industrial goods to large multinational companies who serve both industrial customers and consumers.

The Strategic Services group occasionally does projects that need the use of survey research data as a part of an overall strategic study. Computer interviewing software has been used on an experimental basis during the last 18 months with a high degree of success. These projects have included consumer financial services, office equipment, and industrial chemicals. The software has been used both to augment and to replace techniques that have been used in the past.

Ci2 (Ci2 System for Computer-Interactive Interviewing) was particularly useful for the financial services research. The interview was with 400 respondents with disposable assets of . 200,000 or more. These respondents were asked as many as 30 (out of more than 50) questions about the size and nature of their current holdings. The responses to these questions were used by our customized module outside Ci2 to determine which of 3 conjoint tasks the respondent would be asked to do and then this information was passed back to Ci2.

This interview would have been impossible to do without the computer. Previous studies by the client had produced a 12 - 15% "refuse to answer" response to many financial questions and the algorithm used to determine the conjoint task was too complex for an interviewer to use in the field.

We speculated (correctly) that respondents would be more willing to answer personal financial questions in a one-on-one interview with a computer than with a person. Indeed, the "refuse to answer" rate was less than 1% for most questions and the administration of the interview was smooth and effortless for the respondent and the interviewer.

Preparing a project for field research abroad is much the same as preparing one for the U.S., only more so: more preparation, more pre-testing, more training, more direct control of the field work and more expensive. These are the major areas that all require more attention:

* Selecting the field organization
* Computers for interviewing
* Questionnaire preparation
* Interviewer Briefing

Field Organizations

Generally, there's less mall intercept interviewing and far more door-to-door interviewing outside the U.S. If respondents are pre-recruited for an interview (either by telephone or by intercept on street corners), it is far more likely that an appointment will be made for the interviewer to come to the home/office of the respondent than for the respondent to be asked to come to a central location.

Unlike the U.S. where field interviewing tends to be done by a network of companies serving a city or a region, field houses abroad tend to service an entire country. This is a because of geographic and/or population density. Many interviewers are independent contractors who may work for multiple field organizations.

Most of these field organizations are a part of full-service research suppliers. Most don't have any experience with computer interviewing. Those that have done computer interviewing have used mini-or mainframe-based CATI (Computer-Aided Telephone Interviewing) systems. Here what we look for in a supplier:

- o Established reputation for quality work

- o Excited about the opportunity to do computer interviewing

- o That we/they feel we can work together in a true partnership to achieve quality work.

The partnership part of the relationship is important. We've found that it's important that we take on some functions that a supplier might normally provide. We view this as a part of helping the supplier in learning to do computer interviews and as being essential to the quality of the field work during the first project with a supplier. We expect to:

Train all the interviewers who will be working on our project

Have direct contact with the interviewers during the field work

Have work sent to us on a daily basis

We also want the supplier to provide more help to us than they might normally. We want them to take part in the language translation process (if needed) and help guide us through the pitfalls of the local culture. Altogether, an open, trusting partnership with shared responsibilities for the success of the entire interviewing process is needed for a successful project.

Computers for Interviewing

We have to provide our own computers for the interviewers to use. This turns out to be an advantage as it gives us a known environment to work in: monitor, DOS version and character set.

We use the Compaq 286 portable III with a 20mb hard drive. This machine is widely available and we're able to rent in quantities of 10 - 40 at competitive prices. The computer has a legible screen and is reasonably portable.

The hard drive is essential. Its speed and capacity are needed to provide a menuing system for the interviewer and to switch from module to module during the interview. It cuts down on the floppy disk handling. The interviewer only needs to make a menu selection and insert the current backup disk at the completion of each interview.

The machines are usually rented from 1 or 2 suppliers. As a part of our contract we arrange for insurance and service for the machines. We've found that it's common for the Compaqs to have problems with faulty floppy drives and batteries for backup memory. We specifically ask that the machines we rent be checked for these problems.

Questionnaire Preparation

The preparation of a questionnaire is a team project. It usually involves working from prior small scale quantitative/qualitative research that may have been done in multiple languages. Questionnaire preparation is never simple.

The translations needed for a multi-language project add considerable effort. The translations, just as the questions, need developing through a process of translating-testing-retranslating. Several iterations are called for.

The questionnaires and translations are developed directly in the computer. We use a flow chart to layout the questionnaire in block format and use this as a guide for writing both the questions and logic. It's handy to have a copy of the DOS "keyb" command handy and a chart of the various keyboard configurations for various languages. Compaq, for example, has more than 10 different keyboard configurations.

The batch files need carefully testing. It's common for us to use Ci2 with 6-9 other modules, using the "OUT" instruction. We use the OUT instruction to branch to different conjoint tasks, invoke a timing module to time the length of different parts of the interview (and time/date stamp the interview and record key parts of the questionnaire for a quick check of the field work) and to invoke custom question modules for questions that are difficult to do in Ci2.

Interviewer Briefing

The interviewers in the field and their respondents are the people doing the field work; the essential part of the project. We want them to "buy into" the project. It's our responsibility to:

- Sell them the project
- Train them in the use of computers (many have never seen one before)
- Train them in using the software as required
- Make them secure in the use of the computer and our questionnaire
- Listen to the interviewers, treat their comments and criticisms with respect, and be willing to make changes to accommodate them.

We prepare a 15-20 page briefing book that's designed to be used during the training sessions and as a reference manual for the interviewer in the field. The briefing book contains:

Overview of project
"Key" questions in Ci2
Samples from each section of ACA System (Adaptive Conjoint Analysis)/APM System (Adaptive Perceptual Mapping)
Summary of Ci2 special keys
Flow chart of interview
Glossary for questionnaire
How to use the computer - basic computer instructions
Glossary of computer terms

The interviewer training session is scheduled for 2 days. We train 10-20 interviewers at a time using 3-5 team members for the training. Here's an outline for a typical training session:

Day 1

Introduction and overview of the training session
Overview of the project
Using the computer
- setting it up / turning it on
- special keys
- using floppies
- the help line

Lunch
A quick run through the interview
Using the specific software
A detailed run through the interview using the computer and flow chart
Open discussion of the interview
Review of computer, floppies, etc.

Day 2

Open discussion of interview, computers
Practice interviews - non-directed
Open discussion
Lunch
Practice interviews with each other
Open discussion
Set up computer for the field

Summary

Computer interviewing works outside the U.S. (Sawtooth Software supports any language that's supported by DOS.) The design/translation of a questionnaire for a study in a multi-language project is the same as it would be for a pencil- and-paper interview. The training of interviewers in the use of a computer is the same as training for any new research technique.

We've found that using computer interviewing increases the range of questions we can expect to get answered, increases the flexibility of our designs allowing us to do things we couldn't do with paper and pencil, and decreases the time required between field work and availability of results.

Research In Japan
A Brief Appetizer

Ray Poynter
Sandpiper International Limited

Sandpiper has been conducting CAI (computer aided interviewing) since 1980, and in Japan since 1983. At present about 5% of our world-wide CAI projects involve commercial software; the remaining projects employing software written in-house.

The Gaijin

The Japanese word for foreigner is gaijin, an "outside person." This simple word will define your entire experience in Japan. The Japanese approach to a gaijin is double-edged, in terms of advantage and limitation.

The Japanese will tend to overlook breaches of code, etiquette and propriety, on the grounds that the foreign businessperson is only a gaijin. Thus, if you make some gaffe, and you are sure to, it will not be held against you, as you are only a gaijin. However, the foreigner will always remain a gaijin, an "outside person." No amount of time or effort will fully integrate a gaijin into Japanese society and culture. When you stand, the only Westerner, on some busy Tokyo street, it is only too easy to comprehend and indeed feel the sense of separation between the Japanese and the gaijin. At best, a long-term resident might hope to become a "strange-gaijin." But while even a strange-gaijin may be accepted by his or her close friends, to the rest of Japan that person will just be a different form of gaijin.

Three types of Business:

Sandpiper business in Japan can be divided into three basic categories:

- Research undertaken in Japan and sold to Western corporations
- Research undertaken in the West and sold to Japanese corporations
- The "tough nut" research undertaken in Japan and sold to Japanese corporations

The latter two categories require a Japanese presence. Sandpiper has had an office in Tokyo, staffed by Japanese nationals, since 1983. However - and this is perhaps the main thrust of this paper - it is the first category which represents the most exciting challenge to any market research organization.

The question is often asked: Why, with all the difficulties, work in Japan? The answer is simple: market forces. Japan has a population of 120 million, making it, in population terms, the world's second-largest developed economy, with only the population of the USA being larger. In economic terms, Japan's GNP is comparable to that of the USA, and with Japan spending less than 2% of its GNP on defense, all this together makes the Japanese market potentially the most inviting spot on earth.

The market opportunity that exists for Western research organizations is to provide an information interface between the Japanese consumer and the Western corporation. Those Western corporations that have tried to work with a Japanese research organization have often found it as confusing as dealing with the Japanese consumer direct. Western corporations with Japanese branches often feel they are insufficiently involved in the design of an appropriate market strategy, and that they are too isolated from any research being conducted.

When considering the Japanese market, it is important to note two specific economic features, one on the demand side and one concerning supply. Firstly, the Japanese consumer is a voracious consumer. High land prices mean that most Japanese could not even consider buying a home. This, combined with Japan's high GNP, has resulted in a nation of highly active consumers.

On the supply side, the Japanese corporation tends to be driven by a different basic motivation as compared to its American counterpart. The typical American analysis will at some stage focus on profit maximization, even if this would result in a loss of market share. The Japanese approach tends to concentrate far more on market share, profit being seen simply as a minimum requirement; something that must be attained but not an end in itself. When looking to set a price for a product that a Japanese company is intending to introduce, they will often reject their current supply curve. Instead, they will estimate what their supply curve would be like once they had corrected production faults, established distribution, developed new production techniques, and so on. With this input they will use the new supply curve to set their price at a level that will return a profit, perhaps one, two, or more years into production.

Collecting Data

Just about the simplest thing to consider in Japan is sampling. The majority of projects in Japan use "mall intercept"-type interviews in just two regions: Keihin and Keihanshin. Keihin is the region that includes Tokyo and Yokohama; Keihanshin is the region of Osaka, Kobe, and Kyoto. The population of Tokyo alone is 12 million, so there is plenty of scope for different sampling scenarios within these two regions.

The three things that make interviewing difficult in Japan are the spoken language, the culture and the character set. The first two problems apply equally to CAI and to pencil-and-paper interviews. The problem of the character set has, hitherto, been unique to CAI.

There are four ways of writing in Japan, all of which are in common usage:

Kanji

Kanji are the ideograms normally associated with both Japan and China. They are in fact of Chinese origin. Each Kanji represents a single, simple word. Each Japanese word is made up of one or more Kanji, but normally no more than three. Kanji are complicated and normally require a 16x16 matrix of pixels to display them on a computer monitor; Roman characters require only 7x9 pixels. To represent normal Japanese, approximately 3,000 different Kanji are required. While the meaning of an ideogram is constant, its pronunciation is not.

Hiragana

Hiragana is a much simpler character set than Kanji. There are about 50 characters plus about 50 variants. These simple and cursive characters each represent a single sound. They allow the Japanese to spell out the sound of a word whose Kanji form they may not instantly be able to bring to mind. Hiragana are sometimes used for simplicity even when the author may know the Kanji perfectly well.

Katakana

Katakana is similar in function to Hiragana in that each Katakana symbol represents a single sound. The characters are more angular and even easier for Western eyes to differentiate. There are about as many Katakana characters as there are Hiragana. Katakana is used to represent Western words such as coffee and Coca-Cola.

Romaji

Romaji is an attempt to represent Japanese words through the Roman alphabet. The most common system of spelling is known as the Hepburn system. This system is used in many public places to spell out Japanese words and place names.

While at first it might be thought possible to use Romaji or Katakana to word a Japanese questionnaire, in practice this does not prove to be the case. Although products can often be rendered in Katakana or Romaji, words or phrases used to describe a product's attributes do not convey the same message when removed from the context of their own character set. A phrase is integrally linked with its ideogrammatic image; change one and you change the other. Furthermore, not all Japanese can adequately read the Romaji script. There are at least two MS-DOS machines that fully support Japanese character sets. Toshiba makes a range of machines that are fairly IBM- compatible and have a Japanese language facility. NEC produces machines with a larger user base, but unfortunately with a lower level of IBM compatibility. It is also worth noting that Epson produces a range of NEC-compatible machines.

The Sawtooth Software products Ci2 (Ci2 System for Computer- Interactive Interviewing) and ACA (Adaptive Conjoint Analysis) have been converted to run in Japanese on the NEC machines by the Japanese engineering firm Kozo Keikaku Engineering Inc. So, there are no longer any hardware or software obstacles to prevent conjoint analysis from being conducted in Japan.

The greatest obstacle you will face is in the design of the project. Sandpiper's policy, when working in Japan for a Western company, is to design the project in English first. This is then translated into Japanese and the questionnaire is refined using focus groups and test interviews. The refinement is needed to ascertain that the questions asked are likely to produce sensible and differentiated answers. The refinement must also be seeking to find conceptual areas that the original design might have missed. Are additional attributes required? Are different usage occasions relevant? Is the competitive array of products different from that of another country?

When the questionnaire is more or less complete in the local language, it should be translated back into English to see if the results are likely to be meaningful to a Westerner.

It is vital that this is done before the real fieldwork begins. There is no point in collecting information in a way that cannot be translated back into a Western language for a Western client.

Etiquette

When you do work in Japan you will need to interact with Japanese people, in their own surroundings. Japan is a country where etiquette and protocol are extremely important. You will be well advised to read all you can about Japanese customs.

As previously mentioned, you will be excused many errors simply because you are a gaijin. However, most people would prefer not to be constantly making social gaffes; also, it makes life difficult for existing gaijins when new people arrive in Japan completely unbriefed as to what is considered normal social and business practice.

Two useful notes for etiquette relate to handshaking and to the use of first names. The average Japanese person finds handshaking physically repulsive. However, when dealing extensively with Western people, the Japanese businessperson becomes accustomed to both the idea and the practice. In your initial dealings with the Japanese you are likely to find bowing uncomfortable. The best compromise is to adopt a gentle handshake, not a normal hearty American grip. Japanese people tend to use first names with only their family, childhood friends, and intimates. When you meet someone called, say, Norio Yamaguchi, then you should address him as Mr. Yamaguchi. If you wanted to be more cosmopolitan, then the correct form of address should be Yamaguchisan.

In general, you should aim to adopt sufficient Japanese habits and customs to appear cosmopolitan, but not so much that it might appear that you were trying to out-Japanese the Japanese. The general opinion of gaijins doing business in Japan is that it is commercially and socially counter-productive to be over-Japanese.

USING PERCEPTUAL MAPPING FOR MARKET-ENTRY DECISIONS

Richard H. Siemer
Dow Chemical Company

I am a practical marketing researcher, and by that I mean one who has less academic training than many of authors in these Proceedings, but instead with experience in the use of these techniques in a commercial environment.

When the study I'll discuss was proposed, one of the questions that was asked was, "Who really needs it in today's environment?" The chemical industry was operating at record profit levels. Oil prices were softening with the end of the Iran/Iraq war, which translated to reduced raw material costs. The economy continued at a rate of slow expansion, inflation was moving along at acceptable levels, and currency exchange rates were favorable for U.S. exports. So when one talks about utilizing marketing research technology that was new to us for better decision making, the question that we had to ask was "Can this favorable situation last?" As we all know, nothing lasts forever.

To appreciate Dow's use of techniques like perceptual mapping and conjoint analysis, you need to understand our approach to strategic planning. We believe strongly that good strategic development is critical to the success of a company like Dow. The book Future Shock speaks of the world as getting smaller and time as moving faster. We find this to be especially true in our markets, which means it is more important than ever to "do it right the first time." We consider four major areas in strategic planning: information about the market, our competitors, the commercial and legal environment, and Dow. The analysis of this information provides the basis for successful strategic synthesis.

The internal information we consider for strategic planning includes corporate strategy, corporate culture, the gap between anticipated grow from existing products and their enhancements versus our growth goals, and Dow's strengths and weaknesses. All of this distills to the basis of a competitive advantage for Dow.

The environment in which we compete is also extremely significant. This is more true today than at any time in the past. Things like legislative hurdles, social impact of our actions, the rate of technological change, and the Environmental Protection Agency have significant impact on our decision making. Gone forever are the days when a company could ignore the broader impact of its decision making. The solid waste issue relative to the plastics industry is a good example of the need to understand and respond to the social and political environment in which we compete. (See exhibit 1).

Now let me borrow a concept from my friend Joe Curry of Sawtooth Software. The continuum of marketing research contribution can go all the way from raw data to information to modelling the future to strategy development. We once talked about translating data into information and the fact that we took great pride in not just delivering data, but in delivering information. But information of this type is usually descriptive and defines the past. What you have to do, though, is to define the future and deal with causal

EXHIBIT 1

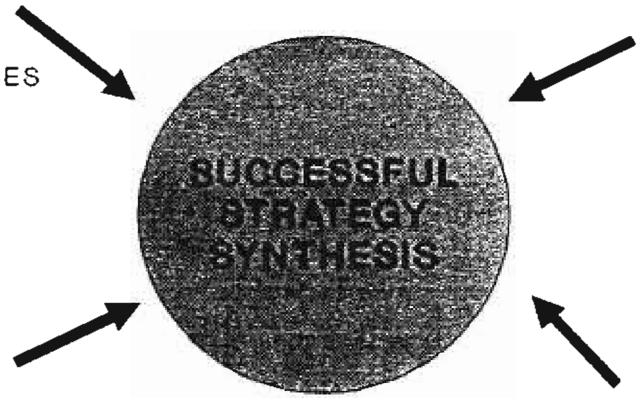
CRITICAL INFORMATION FOR STRATEGY DEVELOPMENT

DOW INTERNAL

- CORP STRATEGY
- CORP CULTURE
- GROWTH GAP
  - STRENGTHS & WEAKNESSES
- BASIS FOR COMPETITIVE ADVANTAGES

COMPETITOR

- STRENGTHS & WEAKNESSES
- RESEARCH LEVEL AND DEPLOYMENT
- VULNERABILITIES
- PROFITABILITY (COST POSITION)
- TECHNOLOGY
- CRITICAL MASS/MKT POSITION
- STRATEGY



MARKET

- MACRO
  - SIZE & GROWTH (BASIS OF GROWTH)
  - PROFITABILITY
  - SUBSTITUTION DYNAMICS
  - SUPPLY/DEMAND OUTLOOK
  - TECHNOLOGY EVOLUTION
  - STAGE OF MATURITY
  - BUYER POWER
  - IMPORT/EXPORT
- CUSTOMER DECISION MODEL
  - NEEDS (REQUIREMENTS,DETERMINANTS)
  - VALUES
  - CULTURE
  - PROFITABILITY
  - INNOVATIVENESS
  - PERCEPTIONS

ENVIRONMENT

- LEGISLATIVE HURDLES
- SOCIAL IMPACT
- RATE OF TECHNICAL INNOVATION
- EPA IMPACT

models. If we change a portion of our marketing mix, how will that impact our position in the market? So we are really interested in causal models that allow us to predict with understanding. It is true that it is more expensive, time consuming, and complex to do this type of marketing research; however, the value far outweighs the investment.

Now let's talk about a recent new product introduction challênge faced by The Dow Chemical Company. Our challenge was rather interesting: How do we enter a late growth stage market against two competent, seemingly well-entrenched competitors? We needed to gain an understanding of all of the information we've been discussing with the objective of defining Dow's best competitive position in the market. That really related to what was important to our customers - the product issues, the service issues, and the price value trade-offs.

Let me describe the market so you can appreciate the situation we faced: We were entering a specialty plastics market in 1984, which was in the late growth phase of its product life cycle. It had been commercialized in the early 1960s and by 1984 was about a $500 million market. The pound consumption growth rate was about three times GNP or approximately 9% a year. The customer base was relatively fragmented, which was new for us. We were accustomed to dealing in large volume plastics where the typically delivery was a 180,000 lbs. rail car. This was a market in which we were dealing with a vast number of smaller customers rather than a few "mega" customers. And we faced tough competition.

Competitor A had about a 72% market share. They were viewed as a marketing powerhouse and had a "take no prisoners" attitude. They also possessed a strong applications development capability. An applications development group works with a customer and shows him how to manufacture a part in plastic that was formerly made from four or five pieces of metal. This type of manufacturing consolidation results in significant savings for the customer. So the market has a great interest in applications development. This competitor also had a broad product line which provided economy of scale and allowed customers to do "one stop shopping." Competitor B had about a 28% market share and excellent product quality. They were technically very competent, and had an extensive European research base as well as a broad product line. (See exhibit 2 for a summary).

On examining this market it seemed as if all reasonable positions were taken. One competitor dominated the marketing side, and the other the technology side. We therefore conducted a study to identify potential positions for Dow Chemical that combined high customer need and competitive weakness. If you think about the strategic issues I discussed earlier, what we were trying to do was to integrate most of those and find an opportunity for Dow.

The study was an extensive one for an industrial department at Dow. We conducted telephone interviews with 402 industry decision makers in 10 market segments. We analyzed the data by advanced statistical techniques including multivariate analysis, correlation analysis, and perceptual mapping. The exhibits show the location of Dow Chemical, competitor "A," competitor "B," the ideal (represented by the star), and the dimensions that distinguish among these entities.

EXHIBIT 2

A SPECIALTY PLASTICS MARKET
1984

- COMMERCIALIZED IN EARLY 1960'S
- SIZE: 296 MM LBS. $600 MM
- GROWTH: 3 x GNP (8-9%)
  (LATE GROWTH PHASE P.L.C.)
- MARKET SHARE

COMPETITOR 1	72%
COMPETITOR 2	28%

- FRAGMENTED MARKET: 6,000 - 10,000 ACCOUNTS

SHEET	31%
ELECTRONIC	25%
APPLIANCE	10%
TRANSPORTATION	8%
SPORTS/RECREATION	5%
PACKAGING	4%
MEDICAL	3%
OTHER	14%

- TOUGH COMPETITORS

1
- MARKETING POWERHOUSE
- "TAKE NO PRISONERS" ATTITUDE
- STRONG APPLICATION DEVELOPMENT
- BROAD ETP LINE

2
- BUYER'S EUROPEAN RESEARCH BASE
- GOOD PRODUCT QUALITY
- TECHNICALLY VERY COMPETENT

Exhibit 3 shows a typical map for a market segment we call "custom molders." With this group, "meeting schedule delivery dates" pointed toward the ideal and therefore the market wanted more of it, whereas "innovation and development" pointed away from the ideal, indicating that Competitor "A" might have been providing this in excess. In total, the market wanted increases in "fair pricing," "meeting scheduled delivery dates," "product consistency," and "support in solving processing problems." These represented opportunities for Dow to gain competitive advantage. Conversely, the market didn't need more of "custom color capability," "innovation/development," "technical literature," or "high heat distortion temperatures."

An important point here is that the market didn't reject these latter issues as unnecessary for a competitor to have in order to succeed. We knew that some of these features were requirements for competitiveness, but they were not going to be the basis for successful competitive advantage.

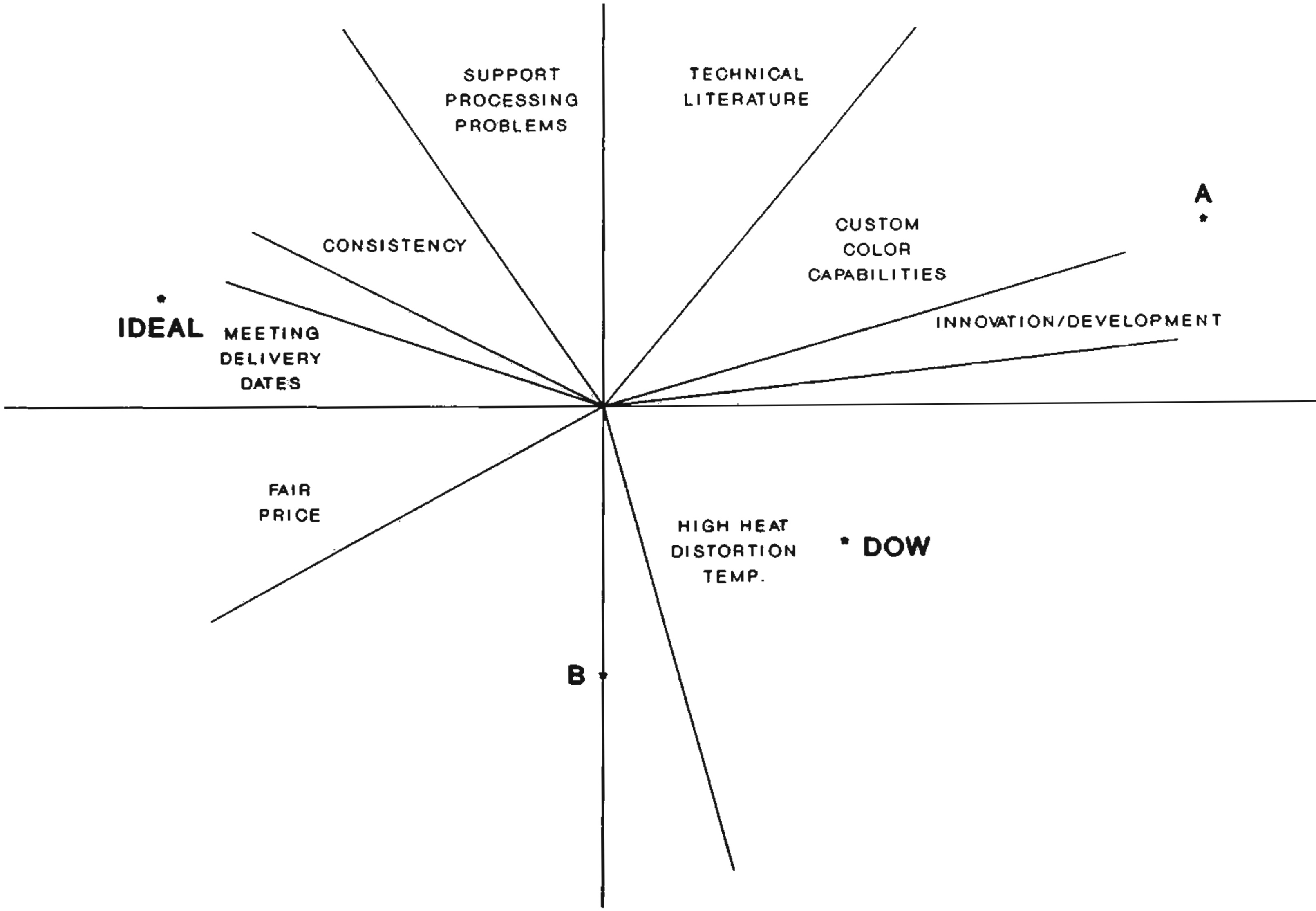
A benefit of this study was the greater understanding of market segmentation compared to our original thinking. As I mentioned, we believed there were eight major industry "need" segments, but we found this was not the case. Instead, we found we could distill the eight into about four major segments. Our original segments 1 and 2 were very similar in their needs and in their view of suppliers. Likewise, segments 4, 5, 6, and 7 again were very similar in their desires, needs, and views of the market. Some of the facts we learned from this study shocked us. Since Dow is a technology company, we had focused on physical product benefits as a basis for competitive advantage. Instead, we found a market more interested in service issues. We found that if we satisfy the two main service issues with distinction, we should be effective in garnering market share. We also found that the primary product benefit will only impact two market segments to a great extent and the second product issue only struck one market.

This study caused changes in our approach to the market and our view of Dow's competitiveness in this market. It also altered our spending plans. So one of the things we learned from this study was greater understanding of the market structure than existed prior to the study. Second, we gained an understanding of the unique needs of the industry segments. Third, understood our competitors' vulnerabilities from the point of view of our customers. Fourth, we were able to develop a strategic positioning for Dow that focused and prioritized our resources where they would have the greatest competitive advantage, and then were able to abandon issues and priorities that had low potential return because of customer indifference. And finally, this study gave direction to product research and development activities.

I discussed earlier the potential of being able to model a market to gain insight into the future. This is the type of output we gained from another study of an industrial market which used conjoint analysis. This study was completed using the ACA System (Adaptive Conjoint Analysis) from Sawtooth Software. Disks were mailed to a group of prescreened respondents who were industry decision makers. Again, our goal was to define Dow's basis for competitive advantage and recommend changes in the marketing mix to achieve this. What you see in exhibit 4 is the impact of a 10% increase in each of the factors on the left, and the expected Dow market share increase in a competitive plastics market that we dominate with a 22% share. If we convince the customer that we deliver 10% more of factor A, our market share would go from 22% to 28.5%. If we are equally effective with factor B, our share will increase from 22% to 26.6%. You can see that the share increase falls off very rapidly after the first three issues. As I mentioned earlier, these are share point

EXHIBIT 3

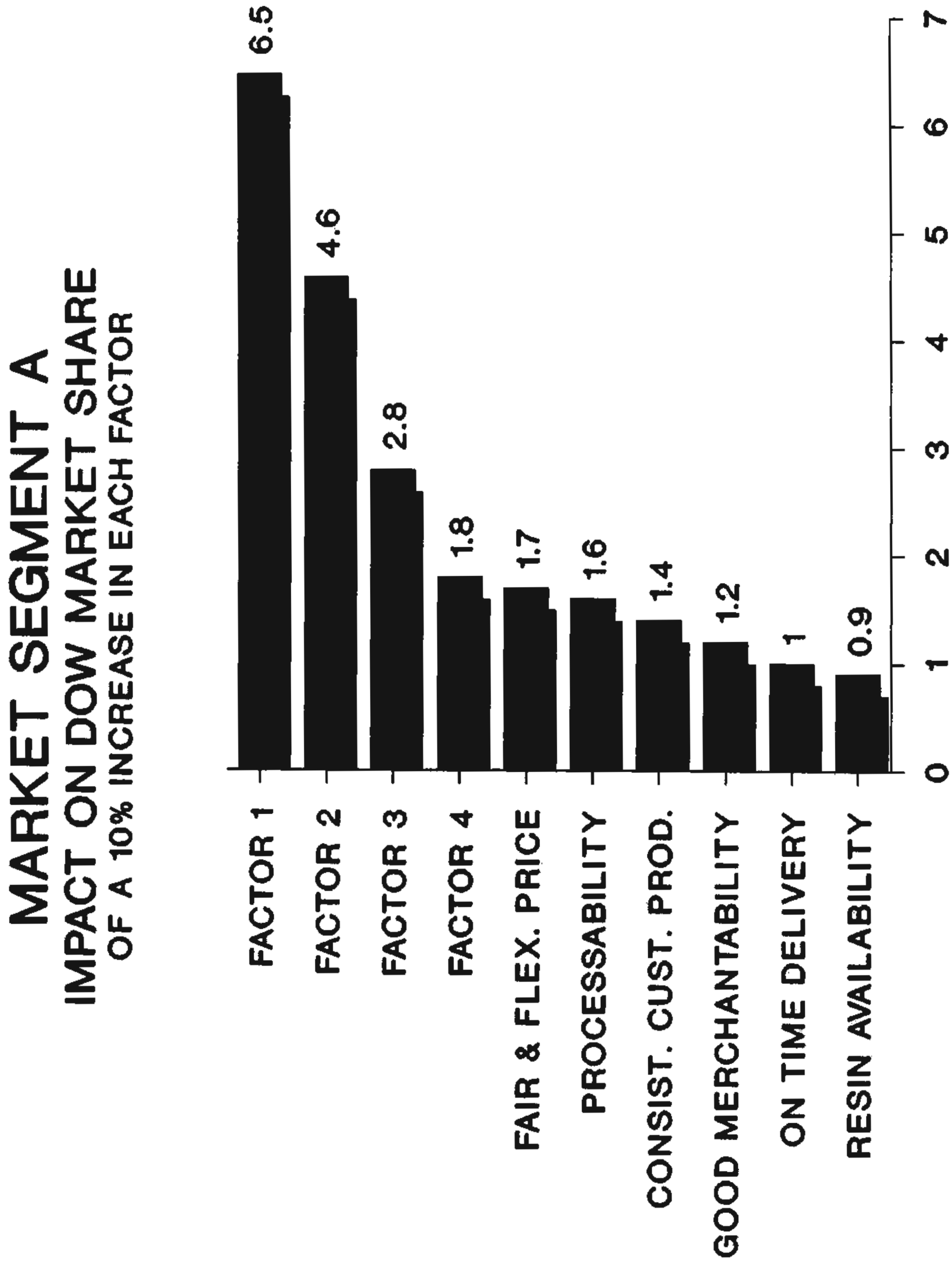
SPECIALTY PLASTICS - CUSTOM MOLDERS
PERCEPTUAL MAP



increases so a boost of 6.5 share points is an astounding change. Factor 1 and factor 2 are closely related. One is the benefit, and the second is the enabling mechanism. Someone might ask if we believe our share will increase exactly 6.5% share points? That exact number is of secondary importance. What is most important for marketing decision making is the rank order and the relative magnitude of difference. These are just two examples of the use of what to us is new technology, which has allowed us to gain greater insight into customer perceptions about Dow and our competitors. We've been able to fine tune our products and marketing programs because we've modeled customer decision making and have been able to understand the impact of changes in our marketing programs. Further, these techniques have the capability of responding to "what if...?" questions that might be raised by our clients after a study is completed. You simply enter the changes and see their impact because you have got a causal, predictive model. It gives you tremendous power.

In the middle ages when man's knowledge of the world beyond his horizon was scarce he frequently assumed the worst. To show this danger when drawing maps, he used symbols of monsters and dragons, and labeled the area "Here Be Dragons." These were places of the unknown and of fear. In the world of marketing there are also those areas labeled "Here Be Dragons." Our job as researchers is to shed light and knowledge on these areas of the unknown, and the tools we've discussed today will certainly help you do that.

EXHIBIT 4



MARKET SEGMENT A
IMPACT ON DOW MARKET SHARE
OF A 10% INCREASE IN EACH FACTOR
FACTOR 1
FACTOR 2
FACTOR 3
FACTOR 4
FAIR & FLEX. PRICE
PROCESSABILITY
CONSIST. CUST. PROD.
GOOD MERCHANTABILITY
ON TIME DELIVERY
RESIN AVAILABILITY
6.5
4.6
2.8
1.8
1.7
1.6
1.4
1.2
1
0.9
0
1
2
3
4
5
6
7

EVALUATING DISTRIBUTION CHANNELS WITH PERCEPTUAL MAPPING

Robert A. Block
John Morton Company

INTRODUCTION

Typically, perceptual mapping is used by marketers as a convenient and effective means of depicting the strengths and weaknesses of product or service offerings relative to those of competitors. The technique provides management with a "snapshot" of how customers perceive their brand(s) to be positioned along the key dimensions used by those customers to differentiate among competing offerings.

I will share an alternative, less typical, application of perceptual mapping that we undertook for an industrial client. We did not seek to evaluate the performance of our client's brand relative to its competitors; rather, we sought in this application to evaluate for our client the suitability of certain channels of distribution to sell its products and service its customers. We then sought through a second, more traditional, application of perceptual mapping to determine how the "suitability positioning" of our client's predominant channel has influenced its brand positioning.

I will first provide you with some background on our client's product offerings and current distribution practices. I will then offer an overview of the complex relationships among manufacturer, distribution channel, and customer. Next, I will discuss how perceptual mapping was used to help our client better understand these complex relationships. And finally, I will show two perceptual maps: one depicts how customers perceive alternative channels of distribution to perform various functions; the second illustrates how our client's current channel practices have translated into its current "brand" positioning.

BACKGROUND

Client Product Portfolio

The client is a manufacturer of electrical equipment and electronic components, products, and systems. Our client seeks to achieve overall company growth primarily through expansion of its high-technology operations. Among this company's primary corporate objectives, therefore, is to be recognized among its customers as a "Technological Leader."

As depicted in Exhibit 1, our client has organized its operations into three major product areas: A, B, and C. Products in area A are electrical/mechanical products, most of which are commodities with a long life cycle--for some, 25 years or more. The majority of sales of these products are to the construction market, generally sold over-the- counter. For the distribution channel, little is required beyond the traditional functions such as stocking, bidding/price negotiation, and delivery; technical training and expertise are not necessary.

EXHIBIT 1

PRODUCT PORTFOLIO

A	B...B	C
Electrical		Electronic
Commodity-like		Highly specific
Mature		Evolving
Long lifecycle		Short lifecycle

MARKETS SERVED

Construction	Industrial

At the other extreme are products in area C, which, for the most part, are electronic and highly specific. These products are rapidly evolving and increasingly more sophisticated, thus leading to swift obsolescence - a much shorter life cycle than products in area A. The majority of sales in product area C are to the industrial market. For the distribution channel to market these products effectively, a large commitment of human and, frequently, capital resources is required. The channel's staff must possess technical expertise and commit to ongoing training in evolving product areas. This staff must, in turn, provide technical assistance to customers, ranging from systems layout to product installation and start-up.

Given the client's current organization, products in area B are quite diverse in terms of the criteria presented in Exhibit 1. In their level of sophistication and degree of differentiation, some are much like those in product area A, others are quite similar to those in product area C, and still others span the broad range in between. As a consequence, marketing these products requires the distribution channel to be quite versatile.

Channels Currently Utilized

Our client currently sells the vast majority of products from each of the three areas through a single channel: electrical distributors. It has long-standing relationships with most of these distributors and, as a manufacturer, its name is well entrenched on their shelves. There exists, however, a variety of alternative channels through which our client might elect to distribute all or certain groups of its products. Among these are:

- Direct salesforce, i.e., sold/delivered directly from the manufacturer to the customer
- Electronics distributors
- Industrial wholesalers
- Manufacturer's representatives
- Original equipment manufacturers (OEM's)
- Systems integrators
- Catalog houses
- DIY/Home centers

Program Objective

Given this background, the objective of our program, as it relates to the application of perceptual mapping, was to assist our client in evaluating the appropriateness of its current single-channel strategy to the needs of its markets. Further, we sought to identify through which channels our client should sell specific products to 1) meet best the unique needs of the markets for those products; and 2) position itself best relative to its competitors within these markets.

Figure 1 provides a simplified overview of the complex set of relationships that operates among manufacturer, channel of distribution, and customer. As represented here, the channels through which the manufacturer distributes--that is, markets--its products act not only as a means of physically moving these goods, but also much like a screen through which the customer perceives the manufacturer. As a consequence, how customers perceive that these selected channels of distribution meet their needs colors how those customers perceive that manufacturer. This, in turn, will influence the manufacturer's competitive positioning.

Premised upon this understanding, we felt that perceptual mapping of channels of distribution would provide our client with valuable information regarding 1) what distribution functions customers need and 2) how suitable they perceive alternative channels are to fulfilling these functions. Because products/markets vary dramatically in what they require of the channel, this analysis would be undertaken separately for product areas A, B, and C, and by customer segment. Combined with the understanding of how our client was positioned among its competitors (derived via perceptual mapping of brands), the mapping of channels would indicate not only how to select channels to better meet customer needs, but also how channel strategies could support the desired competitive positioning.

Exhibit 2 summarizes why we felt perceptual mapping was a tool appropriate to supporting channel selection decisions. In terms of its focus, the input required, and the output desired, we saw that this application had strong parallels to the more traditional brand-oriented application.

APPLICATION

Channel Selection Factors

As part of a computer-administered interview, customers were asked to rate a logically determined subset of distribution channels on 30 functional items. These items covered the entire selling process, ranging from "Makes personal sales calls" to "Delivers when promised" to "Offers post-sale training." Channels were rated on these items along a five-point scale from "Provides very well" to "Does not provide at all." In addition, customers were asked to indicate the importance of each of these functions to them in selecting the channel through which to purchase.

Throughout both of these exercises, customers were asked to focus only on their purchases in one of product areas A, B, or C.

Using principal components analysis based upon channel ratings, the 30 items were reduced to the following seven channel selection factors:

- Product Delivery
- Price
- Sales Assistance
- Ease of Doing Business
- Product Accessibility
- Technical Assistance
- Auxiliary Services

Figure 1

OVERVIEW

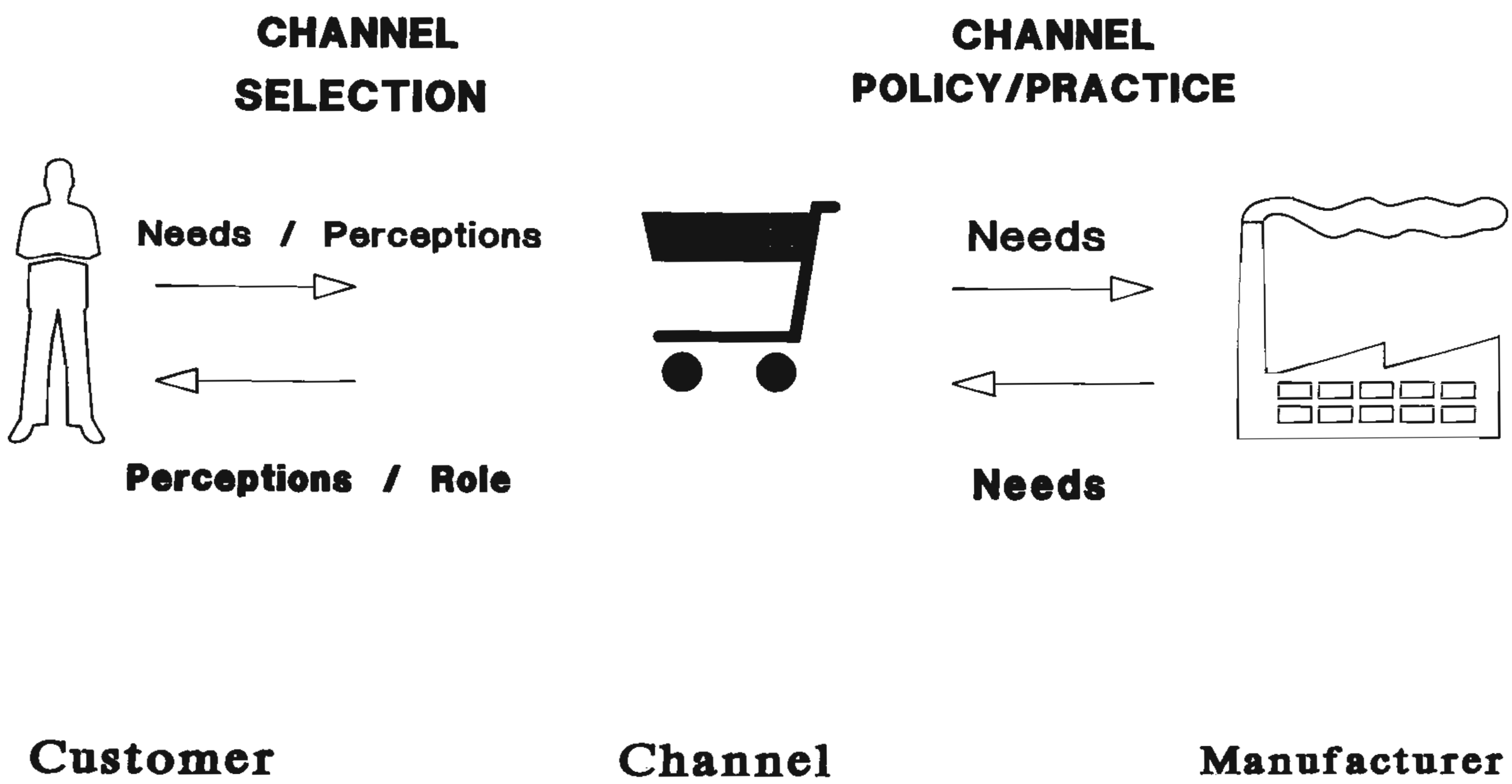


EXHIBIT 2

WHY DID WE USE PERCEPTUAL MAPPING?

	Traditional	This Application
FOCUS:	Relative performance/image of competitors	Suitability of channels to perform specific functions
INPUT:	Ratings on attributes/benefits	Ratings of performance in functional areas
OUTPUT:	Underlying perceptual dimensions: key discriminators among products	"Functional" dimensions
	Positioning of brands along dimensions: image	Perceived ability of various channels to deliver along functional dimensions
	Positioning strategies/general product guidance for client product(s)	Channel selection strategies: mix/match channels to functions

Individual customer scores on each of these factors were then input to discriminant analysis to derive Map 1.

Perceptual Map of Distribution Channels

As Map 1 illustrates, the channel through which our client distributes the majority of its products (electrical distributors) is perceived to provide best "product accessibility," "product delivery," "ease of doing business," and "price." On ,the other hand, electrical distributors are not currently perceived as a credible source of "technical assistance" and "sales assistance." Given our client's objectives of being a technological leader and of achieving growth through the high-technology area, this channel weakness is of strategic importance.

Conversely, the manufacturer's direct salesforce, systems integrators, OEM's, and manufacturer's reps are viewed as credible sources for "technical assistance" and "sales assistance" but are not believed to offer "accessibility" and "delivery."

(Catalog houses and DIY/Home centers are not perceived to offer much more than adequate "accessibility" and thus were dismissed as serious options around which to build a channel strategy.)

An analysis of the perceptual mapping data by product area and market segment revealed that while perceptions were consistent across subgroups, the importance of channel selection factors varied significantly. Specifically, we found that when purchasing products from area A, customers regard "product accessibility" and "product delivery" as more important. This suggested that for marketing these products, electrical distributors were well suited. When purchasing products from area C, however, customers place much greater importance on "technical assistance" and "sales assistance." This, of course, suggested that for our client to further its objective of technological leadership, it should market these products either directly to customers or better utilize alternative channels (for example, OEM's or systems integrators). Alternatively, our client might invest in programs to enhance the technical capabilities of electrical distributors. This might ultimately alter the perceptual space.

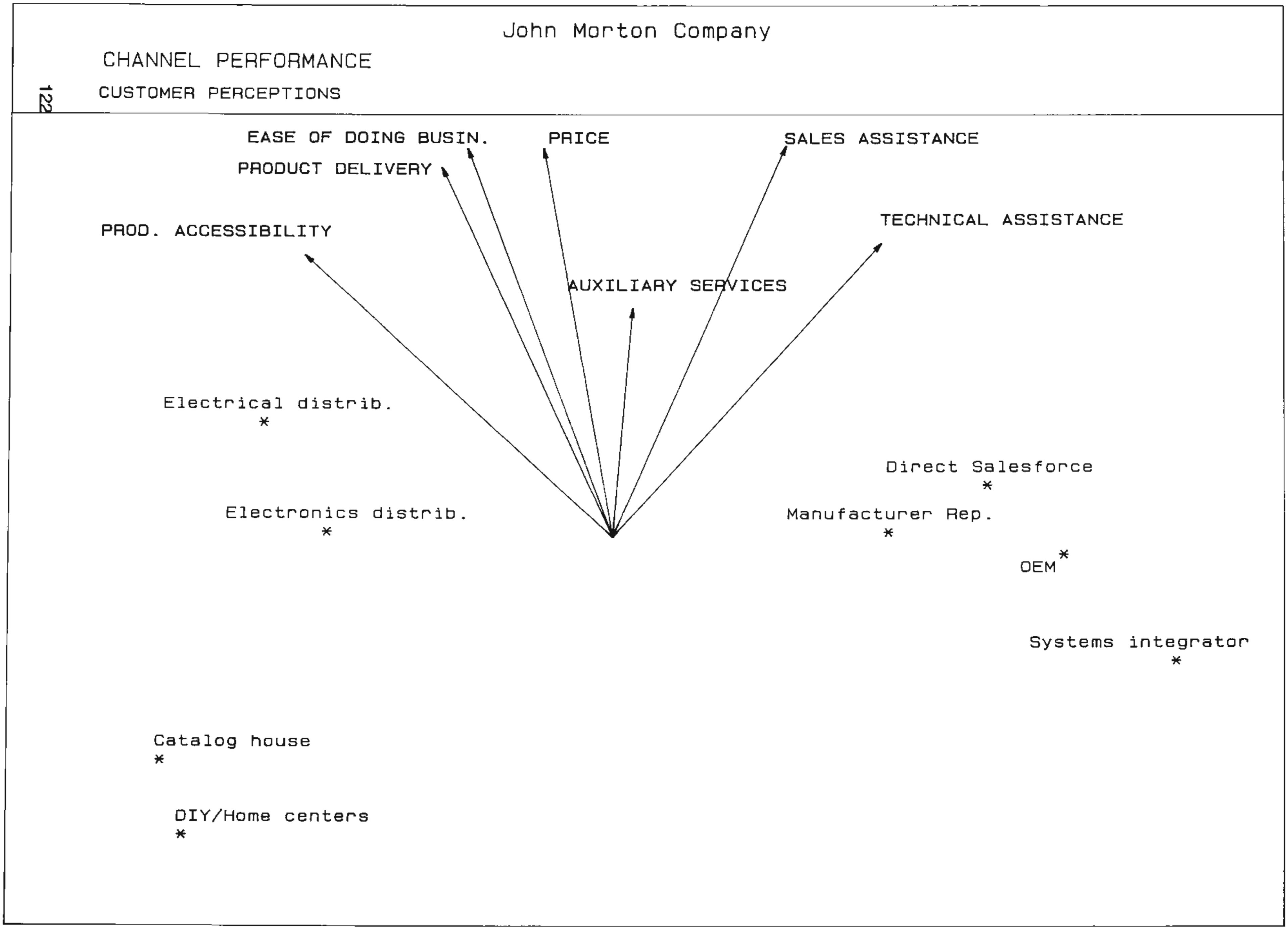
Product area B is less clear-cut. Products in this group are so diverse that "accessibility" and "delivery" are paramount for some while "technical assistance" is essential for others. This might suggest re-organizing to better align operations with market and channel requirements.

Perceptual Map of Client "Brand" versus Competitors

As we discussed earlier (Figure 1), the channels through which a manufacturer markets its products may exert a strong influence on customers' perceptions of that manufacturer. This assertion is corroborated for this market in Map 2, which depicts our client's position relative to its competitors in the market for product area C.

Accept, if you will, that our client's products are considered by experts to be technically equal or superior to those of its major competitors. As you will recall, our client markets these technically intensive products through electrical distributors, which are not perceived to be technically proficient. Competitor #1, however, has established a strong, technically

Map 1



John Morton Company
CHANNEL PERFORMANCE
CUSTOMER PERCEPTIONS
EASE OF DOING BUSIN.
PRICE
SALES ASSISTANCE
PRODUCT DELIVERY
TECHNICAL ASSISTANCE
PROD. ACCESSIBILITY
AUXILIARY SERVICES
Electrical distrib.
Direct Salesforce
Electronics distrib.
Manufacturer Rep.
OEM
Systems integrator
Catalog house
DIY/Home centers

capable network of direct salespeople, systems integrators, and distributors who specialize in its products. As a result, our client is perceived to be much less of a technological leader and has a weaker brand image than Competitor #1.

CONCLUSION

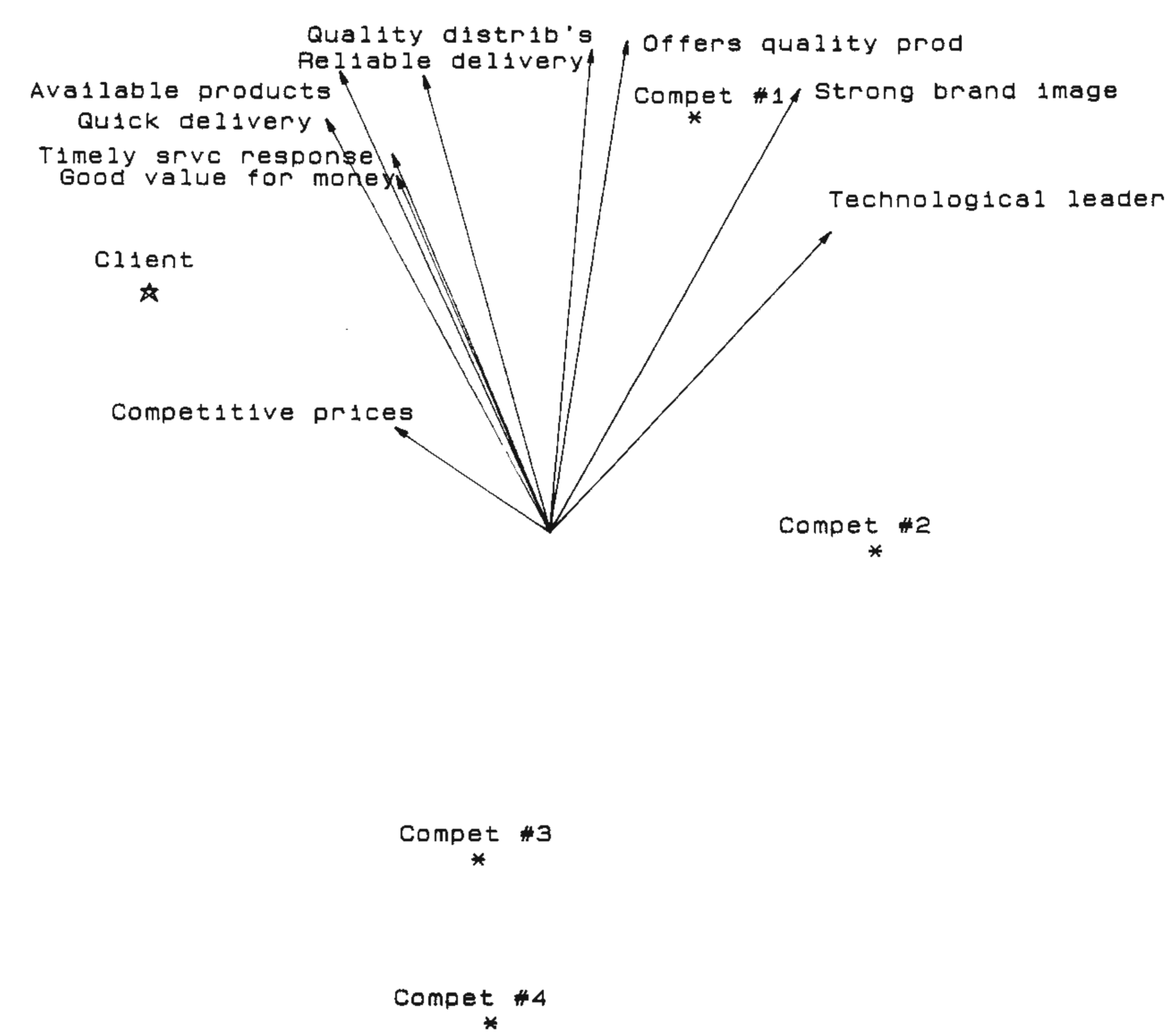
I present this case not so much to emphasize the results of this particular application, but to underline the versatility of perceptual mapping as a planning tool. While we "perceive" this technique almost exclusively as a means of evaluating the relative performance/image of products and services, we should not overlook its robustness across a variety of applications. For fear of extending a technique too far, let us not unduly limit its creative application.

Map 2

John Morton Company

CLIENT'S EXISTING DELIVERY SYSTEM

MARKET PERCEPTIONS - PROCUCT AREA C



REPOSITIONING A SERVICE:
Advanced Marketing Research and Associated Management Issues.

David L. "Dave" Masterson
Masterson & Associates

INTRODUCTION

From time to time, management needs to use research to assess the market position of a service. This is especially true for a complex technical service offered to a market utilizing protracted buy/sell cycles and involved decision making processes. This paper discusses a research study which combined correspondence analysis, conjoint analysis, and cluster analysis to help position a commercial service. This paper also discusses the management risks associated with strategic market research, as learned and reinforced from this study.

BUSINESS SITUATION

The service being evaluated here is a typical commercial service, being sold and provided business-to-business. It is typically "technical" in that it has many different attributes, various pricing schemes, and requires customer change to implement. The service is sold through a proactive commercial sales force, with sales cycles spanning three months to thirty-six months. Customer-supplier relationships can exist for years.

Like so many commercial services, the one under study here was positioned in an increasingly competitive market, experiencing price pressures and margin shrinkage, and subjected to changing customer expectations. For the first six years of this service, management had tried to position it "at the high end" of the scale, differentiating it with a premium level of service and charging correspondingly premium prices. But competitors had made inroads on the differentiating competitive attributes and customers were displaying increasing resistance to "premium" pricing.

The market research project under study here was initiated by senior management to determine how this service was positioned in the "eyes" of the marketplace and to make strategic decisions on how to improve the operational and financial performance of this service. This research project acquired answers for this senior management team. It also yielded insights on management's risks associated with advanced market research and the importance of experiential judgment in the application of these techniques.

RESEARCH METHODOLOGY

The selection of research methodology for this study was driven by the nature of a commercial service and the informational needs of management. The nature of the service presented several research issues.

Service (product) features affecting research:

* Over 40 attributes/features to the service
* A dozen competitors
* Multiple decision influencers in customer accounts
* Importance of vendor institutional image
* Several components in price/cost

Management's needs for information also presented several research issues in the study.

Management's questions:

* Which service features are most important in the buying decision?
* How are trade-offs made?
* How are we positioned relative to these important attributes/features?
* How are we positioned relative to competition?
* How is the market segmented?
* What are pricing elasticities?
* How is our sales force perceived?

Because of the service features and management questions involved, the research project was broken into a two-phase design. The project combined correspondence analysis, conjoint analysis, and cluster analysis to help make strategic decisions concerning this service.

INTERNAL PREPARATION

This research was probably going to yield some "bad news" to one or more internal constituencies regardless of the results of the research. Each constituency had its own perception of the service's strengths and weaknesses, as well as the relative importance of each functional area of the organization. Several disagreed strongly. Thus this research project was considerably at risk from a possible "shoot the messenger" mentality right from the start. Additionally the management team involved had little to no experience with the research methodologies called for in this project. Indeed, most of management's prior experience with research centered around syndicated studies and simple rating scales. The manager in charge of this research project took several measures to deal with these internal issues and to prepare the various constituencies for the pending results. The first of these measures was to acquire agreement for the need for the research from the General Manager for this service. This agreement was attained and tested by isolating this research project as a single line item in the budget and having it survive through the competing-for-resources phase of the company's budgeting cycle. Thus, not only was "agreement" attained for this project, but also a prioritization and an "importance" were associated with it.

The second measure taken to prepare internal employees and managers was to use an internal committee of functional experts to help define issues and to agree on attribute terminology to be used in the interview questionnaire itself. Like many commercial services, the service under study had many terminology issues and nuances. Various functional areas within this company and within respondent institutions had different connotations for various terms. The manager for this project had to come to internal consensus on meanings and then ensure that respondents were interpreting the terms in the same way. This was a lengthy process. This committee arrangement created a communication forum for inter-functional issues and an opportunity for eight to ten mid-level managers to understand the nature of the research project.

The third measure taken was to insulate the rest of the organization from the nuts and bolts of the project. The senior management team knew of the project's progress, the functional experts were involved in issues and terminology nuances, but the majority of the organization had little, if any, knowledge of the research project. This was done to keep anticipated anxiety levels as low as possible and to minimize the Hawthorne effect on sales and service levels.

The fourth measure affected the design of the research itself. The project was designed in two phases. This two-phase approach accommodated the technical demands of the research, but it also accommodated management readiness issues toward the research methodologies being used. Senior management could learn from, and deal with, Phase I methods and results prior to being exposed to Phase II methods and results. The design also called for relatively large sample sizes in anticipation of what was perceived as inevitable credibility and "believability" challenges.

RESEARCH PROCESS / EXECUTION

As mentioned previously, this research was divided into two phases. Phase I used correspondence analysis for perceptual mapping to confirm buyer decision making processes and perceived positioning of the service. Phase II used adaptive conjoint analysis and perceptual mapping to analyze and make attribute trade-offs and to determine "ideal points" for various market segments.

Both phases used telephone interviewing. In Phase I, eight hundred (800) daytime interviews were completed. Seven hundred fifteen (715) were selected randomly from the total market. Eighty-five (85) were selected randomly from the company's customer list. Initial respondents were qualified as the person within the business most responsible for deciding which service to use. The questionnaire was designed around five functional areas within the customer's business. In this research, the respondent company was the survey unit, not the responding person. After the respondent was screened as being the person most responsible overall for the decision as to which service provider to use, he/she was then screened as to whether he/she was most responsible for each of five functional departments or areas. If not, the interviewer then switched to the person who was, and completed the portion of the interview appropriate for that function. In some respondent accounts where functional sophistication was shared by several decision makers/influencers, a fourth and fifth interview were conducted. In Phase II, six hundred seven (607) daytime interviews were completed. All of these were completed with respondents from Phase I. As in Phase I, all respondents were again qualified as the person most responsible for the purchasing decision. Sixty-five (65) were completed with customers.

Correspondence analysis was selected in favor of factor analysis and discriminant analysis. Since correspondence analysis allowed the respondent to "pick the best provider at...," this process was easier for the respondent, and was more cost effective for the end user. It also allowed for superimposing "ideal" (supplemental) points after the fact. Correspondence analysis was also preferred because of the many service attributes under study and the many competitors under study. Using rating scale techniques covering all these attributes across all the competitors would have been quite difficult. Also, any one respondent may not have been familiar with each competitor or even each attribute.

Eventually, adaptive conjoint analysis was selected for Phase II. This was only after two other conjoint methodologies were eliminated, however. A version of full profile conjoint was tested first. Despite conventional wisdom, this study tried to use telephone interviewing with a "special" version of full profile conjoint. We mailed various profiles to the respondents which had only limited portions of service profiles. We then added more attributes to the profiles over the telephone after screening the respondent, and then asked the respondent to select their preferences. Management had several reasons for trying this. Respondents were dispersed geographically and face-to-face interviewing was impractical. Using a mail survey instrument represented strategic risk, because this type of research had never been done before in this industry segment. Mail conjoint survey materials containing complete profiles would reveal the study design and possibly dissipate competitive advantages in the research results. The additions and changes to profiles over the telephone protected the design because it was virtually impossible for the respondent to reconstruct the profile comparisons after the interview. The full-profile pilot test experienced difficulty and the results turned out to be little more than statistical noise. An incomplete block design was considered second, but was eliminated without testing. The effective sample sizes were going to be too small and management wanted to use simulation as part of the analysis.

Thus adaptive conjoint was selected and Sawtooth's ACA software was suggested for the telephone interviewing. The interview design using ACA was pilot tested with the same group of respondents used to test the full profile design. This time the results were quite satisfactory statistically. Also, informal interviews were conducted with the test group respondents, and they clearly preferred the interview process presented by the adaptive conjoint design; they could understand the profiles and questions better because they were dealing with the attributes and trade-offs which were important to them. Phase II field work was then completed re-interviewing willing respondents from Phase I. Since these second phase data were generated from the same respondents as the first phase data, Phase II data were then appended to the Phase I data. Cluster analysis was done and cluster points were treated as supplemental points on the perceptual maps.

Study results were conveyed to management via management briefings, written management summaries, and formal analytical reports.

BUSINESS RESULTS

This research carried quite a bit of "news" for the senior management team. This company's service turned out to be clearly differentiated and perceived as competitively superior to a specific market segment. Unfortunately, that segment turned out to be much smaller than previously assumed, and even within the segment, customers were making considerable use of competitors' service as well. Additionally, the differentiation and position held by this service almost precluded it from exploiting more profitable segments of the market. Finally, major marketing and sales weaknesses were identified, thus confirming suspicions of several members of the management team. As so often happens, the market held misperceptions about the service and was ignorant about it. Management also had misperceptions about the position of its service within the market.

As of this writing, senior management is taking steps to strategically change the service anc to reposition it within its market.

LESSONS LEARNED OR REINFORCED

This research project proved quite fruitful in terms of lessons learned for both the researchers and end-user management.

Some lessons applied to both researchers and management. The first of these centered around the use of two major methodologies in a single study. Here, the use of perceptual mapping for positioning and conjoint analysis for demand graphs yielded considerable insights into strategic issues. The two methods highlighted different perspectives and richly broadened management's insight into what needed to be done. In retrospect, the results from combining these two methods were far superior to having used only one of them.

The second lesson for both researchers and managers centers around the "people" limitations of respondents, regardless of the sophistication of the methodology. Here, a portion of the conjoint work used a pricing component which seemed quite straightforward and reflected actual market conditions existing a few years prior to this research. Surprisingly, a very high percentage of respondents could not deal with the trade-offs on this pricing component. A majority of respondents either "pleaded the fifth amendment" or simply refused to participate "under that scenario!" Those who did deal with this component generated statistically insignificant results. Thus, management failed to get one piece of desired information simply because of psychological barriers on the part of the respondents.

This research also yielded some lessons which were primarily applicable to the researchers. The first of these is "old news" but it is very important none the less: Complex questionnaire designs involved in complex services need to be executed by competent interviewers. Presenting trade-off options with specialized industry jargon can be difficult for both respondent and interviewer. In this study, interviewers were selected for their competency in industrial interviewing and the company's project manager spent time explaining the service to them. He also spent time discussing some of the jargon associated with this service and explained some of the nuances in the different terminology. This interviewing competency and time investment in training proved very important in the execution of this research.

The second research lesson in this project is in the limitations of conjoint research for service industries. There are so many features to a service, with so many attributes of service manner when it is delivered, that the number of possible combinations is huge. Full profile did not work in this study primarily because it was delivered over the telephone. But even as a mail design was being considered, the demands made upon the potential respondent by a full profile design were staggering. This study reaffirmed the need for adaptive methodologies for services research, while recognizing the technical debates surrounding the adaptive approaches.

The third research lesson highlighted by this effort was the need for flexibility by researchers not to do "perfect" research. This study used more interviews than necessary for reliable results. It prohibited doing conjoint by mail. It knowingly and proactively skewed the design away from one competitor in the market. These were all done for management cultural and readiness reasons; not for research reasons. The adaptability and understanding of the research project manager in these matters was very important. The technical professional's ability to accommodate these business needs contributed significantly to the success of this project.

The fourth lesson for researchers reinforced by this study is not to make assumptions and decisions for the end user. When conjoint analysis is used or when modeling is done, researchers frequently make recommendations directed toward reconfiguring the service offering in order to increase market share or "share of preference." In this study, the company's differentiation was already more than adequate to attract market share. Management's use of the conjoint and modeling was to test service configurations which would improve profit margin and which would relieve various operational pressures which were causing quality problems and morale problems. Management was not particularly interested in market share. Researchers seldom understand what line management is really trying to accomplish via research. Even if they are told, they frequently receive incomplete, filtered or "politically sanitized" information. Assuming management's direction based upon de facto marketing logic can result in "naive" recommendations which can diminish the perceived value of the research.

In some ways, the management lessons reinforced by this project are more broad and more important than the technical lessons.

The first of these management lesson involves the very nature of doing research for service industries. The sale and delivery of a service is very people intensive. Thus, research on a service can be quite emotional and results can be taken personally by the employees and managers involved. As experienced in this study, research projects for services need to be pre-sold well internally, contain relatively large and "believable" sample sizes, and deliver results in controlled quantities that managers can digest comfortably. Even under these circumstances, resistance to research findings can be high, especially in the face of emotional responses to "bad news."

The second management lesson reinforced in this study involves management readiness for so-called "advanced" methodologies. Because of the technical difficulties inherent in multi-attribute services research, the application of these methodologies is a fairly recent phenomenon for many service managers. Thus, managers are frequently more conditioned to, and more comfortable with, simpler ordinal rankings and "importance" scales. When advanced research methods present "bad news," management's lack of experience with complex reporting methods can trigger challenges and credibility battles. This can obviate any benefit in the research investment and can affect the proverbial "research sitting on a shelf." The lesson here is to ensure that perceptual maps cover a high percentage of the data variability, to complete and report the research in phases, and to patiently but effectively explain how to interpret results.

The third management lesson reinforced in this study is the criticality of having competent experience and judgment positioned between the company's professional line management and the research professionals designing and executing the project. Line managers are seldom statistically oriented and researchers will seldom understand the political and cultural obstacles to implementing change in a service organization. Successfully navigating the people intensity of a service, the many attributes involved in services, management readiness with advanced methods and the technical pros and cons of adaptive research can be quite challenging. Management needs recommendations that are real-world "doable." Researchers need clear direction and timely decisions at a project's crossroads. Applying advanced methods on behalf of a service provider expands the traditional need for competent perspective sitting on the fence and looking into both backyards.

A fourth, and somewhat related, management lesson is in the gap between statistical market segmentation and "sales-force-doable" segmentation for services that primarily use a sales force for their communications vehicle. In this study the cluster analysis resulted in four fairly distinct market segments with clear preferences and different service needs. Unfortunately, these differences were primarily psychographic, not demographic. To exploit these differences and better meet the needs of these segments, the company's sales force would need to be competent at a fairly sophisticated sales methodology. In this study, as with many positioning efforts delivered primarily via a sales force, success and effectiveness in segmentation was going to depend on the sales force's ability to execute it, not on the research's ability to identify it. As reinforced here, advanced research methods can identify opportunities of which the organization simply cannot take advantage in the short run.

Thus this research effort yielded several valuable lessons. Two applied to both researchers and end users, four applied primarily to the researchers, and four were more applicable for user managers.

Perceptual Mapping and Cluster Analysis: Some Problems and Solutions

Charles I. Stannard
D'Arcy Masius Benton & Bowles

This paper will discuss some common problems and issues analysts have to deal with in studies using perceptual mapping and cluster analysis. It will describe the problems and the various ways we dealt with each of them in specific studies. The first problem, common to much marketing and advertising research, is how to deal with owners and nonowners of a brand. The context for this discussion is a large study of the appliance category. The second issue concerns evaluating segments based on cluster analysis. The data are also from the appliance study. The third issue concerns the use and interpretation of maps in advertising research. Here the data come from a study of the automotive category.

Dealing with Owners and Nonowners

The context for the discussion of owners and nonowners is a large positioning study conducted for a major maker of appliances. The study was done a year ago for the purpose of providing direction for the development of image objectives for the brand. We wanted to understand the characteristics (attributes and benefits) by which purchasers of major appliances distinguish manufacturers, and to determine the importance of these characteristics in the purchase decision. The project had two phases: a qualitative phase to learn which attributes and benefits consumers use to differentiate among manufacturers, and to understand qualitatively the process by which consumers purchase major appliances; a quantitative phase in which we quantified and tested what we learned in the qualitative phase.

Anytime we seek information from consumers the problem of to whom do we talk confronts us. This problem is very similar to the problem confronting anthropologists studying a strange culture. In anthropology it is called the problem of the informed informant. Whether we are anthropologists in New Guinea or market researchers in the United States the problem is the same: while just about everyone we question will respond with an answer, not all "answers" are equally valid and valuable. Naturally we want good answers. But since the canons of objectivity, not to mention feasibility, prevent us from ruling on the "goodness" of each and every answer, we move from evaluating answers to evaluating "answerers." In choosing to whom we talk , we estimate whether it is reasonable to expect that a given person will give provide us with good answers. The key criterion we use in judging whether a person can give us good answers is whether he or she is likely to be knowledgeable about the subject we are investigating. Typically, we make these judgments based on whether the person can indicate experience with the subject in which we are interested. An important indicator of experience is ownership or use of specific products or brands. We codify the criteria for judging the likely value of a respondent in the qualifications he or she has to meet to enter the study.

The major appliance category (refrigerators, ovens and ranges, dishwashers, clothes washer, dryers, and microwaves) has several characteristics that determined the requirements people had to meet to participate in the study. The first characteristic is that the re-purchase cycle is quite long, 10 or more years. Moving and remodelling can shorten the cycle, but typically consumers are only sporadically in the market, usually after a long absence. Second, while consumers frequently have several different brands of appliances in their homes, they often cannot list them by brand when asked in an interview. Finally, except for moving or remodelling, it appears that many purchases are unanticipated, being a quick response to the actual or expected failure of the product.

These characteristics indicate that most consumers have little current knowledge about manufacturers and brands. They become interested in the category when they are about to purchase one or more appliances. Sometimes, as in remodelling and moving, the purchase is foreseen and the search for information about products and manufacturers is somewhat leisurely and thoughtful. In other instances, when a current product fails, the search process is much more hurried and even haphazard. In either case, we think people move from a state of relatively low awareness and knowledge of manufacturers and their product offerings to one of relatively high awareness and knowledge in a short period of time, which is characterized by a comparatively vigorous search for brand and product knowledge.

For our purposes, therefore, we wanted a sample of people who would have greater than average knowledge of and involvement in the appliance category. They would better represent those people who are in the market for an appliance and thus would provide a better picture of the market from their point of view. Therefore, in addition to the usual demographic, appliance and brand ownership qualifications, we wanted people who were recently in the market for a major appliance, or anticipated being in the market in the near future. They best represented the state of mind and knowledge of consumers at the time of purchase and are the target of the advertising and marketing efforts.

In addition to choosing the right people, we also have to ask them the right questions. In a positioning study, we typically ask respondents to do two things. First, we have them tell us what attributes are important in distinguishing between brands. Then we have them rate brands on the attributes. Choosing the right questions means asking them to rate brands they are familiar with on attributes that are important to them in choosing between brands in the purchase decision. Since the number of attributes and brands of interest was too large - 7 brands and 28 attributes - for any one person to rate all combinations of brands and attributes, we had each respondent rate 4 brands on 12 attributes. The brands and attributes were selected as follows: Each person rated the client's brand and three other brands with which he or she was familiar. The attributes were classified a priori into six categories based on their content. Each person rated who attributes in each category. The specific attributes were the two he or she rated most highly in each of the six categories.

The major part of the analysis of the appliance market involved producing a map of the market which located brands and attributes. The advantages of maps are well known. They provide an economical summary of a great deal of data on brands and attributes, in our case 7 brands and 28 attributes. Another advantage of maps is that the audience, usually managers, often finds them easier to understand, more revealing, and certainly more interesting than other ways of presenting the same data. From the analyst's and presenter's points of view, maps are often easier to present and interpret for an audience than are complex tables of numbers and coefficients.

Having thoroughly considered, or so we thought, the important issue of product and brand ownership, we were chagrined to discover that our initial map did not make a great deal of sense. In analyzing it we found much less discrimination among brands than we expected, and what appeared to be some odd juxtapositions of brands. We found some of the large, middle-range brands like Sears and GE were positioned very close to smaller expensive and high-quality brands like KitchenAid and Maytag. Everything we knew about the market suggested that consumers perceive KitchenAid and Maytag as different from brands like GE, Whirlpool and Frigidaire. Thus, instead of shouting "Eureka!," we invoked the first rule of nonsensical analysis: whenever we find something truly new and unexpected in an analysis, look for an error - either in the logic of the analysis or in the data themselves. We know from experience that the odds favoring an error are much greater than those favoring the discovery of something truly new.

We identified two related aspects of brand ownership as possibly causing the strange map. First, the proportion of the sample owning specific brands varied greatly, mirroring the reality of the marketplace. For instance, three times as many people owned a Sears appliance as owned a Maytag, and five times as many owned a GE appliance as owned a KitchenAid. Second, as is usually the case, people rated more highly the appliance brands they owned than the brands they did not own. In fact, the differences between brand owners and nonowners were greater in many instances than the differences among brands, when ownership was controlled. In combination, these two aspects of brand ownership in our sample could be the reason the larger brands ended up in close proximity to some of the smaller more expensive and higher-quality brands.

The obvious solution, if these were the cause of the problem, was to separate owners and nonowners. We did this by creating fourteen brands, seven as seen by owners and seven as seen by nonowners and estimated the space using fourteen brands - the seven original owner brands plus the seven nonowner brands. This approach has the advantage of using all the information (i.e., the total sample of ratings) in the sample, rather than a portion of it, as would be the case if the space were created using only owners. The disadvantage is that it can be difficult to create mutually exclusive and exhaustive groups of owners and nonowners when there is extensive multiple ownership of brands.

We then re-estimated the space, this time using the fourteen brands. The first function or dimension captured the differences between owners and nonowners. It grouped at one end the owners of the various brands and placed the nonowners of the brands at the other end. Furthermore, there was no overlap between brand owners and nonowners of the brands. In effect, the first dimension accounted for the effects of ownership on the brand ratings.

We based our map on the next two dimensions, which successfully described the marketplace. One dimension was price/value. Brands at one end of the dimension were characterized as offering the lowest prices for comparably featured appliances; brands at the other end of the dimension were seen as saving money in the long run. The third dimension described quality in two different ways. One was called "promised quality." Brands offering promised quality were highly recommended by others and promised to honor warranty claims without hassle or difficulty. The other end of the third dimension was "experienced quality." Brands offering experienced quality were seen as a pleasure to own and extremely durable and long lasting.

Thus, we eliminated the negative effects of brand ownership by creating brands to represent nonowners, estimating the space with the owner and nonowner brands and discarding the first dimension. The reason this worked was due to the fact that owners rate their brands higher than they rate brands they do not own, and these differences, in addition to being consistent, are also substantial, generally being greater than the differences among brands. This explains why it was the first dimension. The fact that in most categories owners rate their brands higher than do nonowners suggests that what we did in the appliance category may have greater utility and generality as a solution to the problem of owners and nonowners.

Segments in Positioning Research

In addition to mapping the usual groups or segments of consumers like users and non-users, men and women, and so on, it is also possible to map segments derived from psychographic data. The latter segments emerge from the data in cluster analysis, as opposed to the former which are predetermined according to explicit criteria. Because the segments in cluster analysis are based on psychographics, they can provide richer and fuller explanations of behavior and market structure. That is, instead of saying, based on the interpretation of a map, people choose brand A because it is low priced and readily available, we can, in the ideal case, elaborate on the reasons people choose Brand A. For example, we might find that the people choosing Brand A really are made up of two different segments, one which is very price sensitive because of low family income and another which has very little interest in the category and therefore opts for the low-priced, convenient brand in this category. Of course, this is the promise of psychographically based segments. In reality, though, we know promises are not always kept.

In our appliance positioning study we also included 12 psychographic statements relating to appliances and shopping in addition to the 28 appliance manufacturer attributes. The psychographic statements included items like "In buying major appliances the reputation of the store is more important than the brand name," "When buying appliances, it pays to buy the best model even though it is more expensive," "It is more important to have good appliances in the home than good furniture."

Clustering the items produced four consumer types. We named them "Flashy Flora and Fred," "Needy Nan and Neil," "Classy Carl and Cristy," and "Apathetic Al and Ann." The demographic and psychographic portraits of the groups appeared to have integrity and make sense. For instance, Needy Nan and Neil, as their name implied had the lowest total family income, with 46 percent of them having total incomes of less than $25,000. Concomitantly, they were the least educated with 40 percent having a high school education or less. They also had the largest families and were the second youngest of the clusters. Their attitudes towards appliances fit their demographics. Nan and Neil were very price sensitive. They wanted to buy the lowest priced appliance from among similar makes and models. At the same time, they had to have appliances that lasted, more so than any of the other clusters. Their extreme price sensitivity created a problem for them. They could not rely on an important indicator of quality and durability, the brand name, to help them choose the best and lowest-priced brand of appliance. As a result they had to look to other sources of information to help them choose among brands. They, more than the other segments, relied on two sources to help them do this. One was _Consumer Reports_. The other was whether they thought the manufacturer was a specialist in kitchen or laundry appliances. They took specialization as an indication of durability, an important attribute in appliances for them.

Classy Carl and Cristy, by way of contrast, were the wealthiest segment. They had the highest household income, 73 percent over $35,000, were the oldest on average, and had the largest homes as indicated by the number of bedrooms and bathrooms. This segment also had the largest number of college graduates of any segment. Their views about appliances were quite different from Needy Nan and Neil's. Classy Carl and Cristy were not very price sensitive. They were the least likely of all segments to look for the least expensive brand of appliance. Rather, they thought it paid to buy the best model appliance, even though it was more expensive. For them, however having appliances that were a pleasure to own was also very important, as were appliances that were easy to clean and keep clean. Their attitudes towards appliances were echoed in their views about their kitchens: they were very proud of them and the way they looked. Indeed, it is likely that Carl and Cristy judged kitchen appliances for their looks as well as for their quality and features. Perhaps because they bought the best appliances, Carl and Cristy, of all the segments, had the most positive attitudes towards appliance makers. This was manifest in their agreement with those statements that implied a willingness of manufacturers to value customers and stand behind their products.

While the portraits that cluster analysis creates can be interesting and plausible, it is important that they relate to product ownership and usage in intuitively meaningful ways. In the appliance category, we expected to find sharp differences among the clusters in brand penetration. For instance, we expected to find penetration of the more expensive brands to be greater for Classy Carl and Cristy than for Needy Nan and Neil. And we expected the opposite penetration for the lower-priced brands.

In fact, however, we did not find the expected pattern of brand penetration among the segments. Instead, we found that brand penetration was relatively flat among the segments. This was puzzling and demanded an explanation. In thinking about the purchase process, however, an explanation of the lack of differential brand penetration suggested itself. The explanation focused on two aspects of the retail side of the appliance business. First, retail sales are increasingly dominated by "power retailers" like Circuit City and Fretter's which continually have sales featuring specific brands. Second, appliances are as much sold as they are bought. Salespeople often receive "spiffs" or special sales inducements above the regular commission from manufacturers for sales of their brand or specific models of their brand. When this occurs, salespeople work hard to steer people toward these brands, with a fair amount of success, according to them. When we take these two aspects of the market into account, the lack of differential brand penetration among segments in brand penetration might reflect a retail reality which is working against the manufacturers' efforts at creating and sustaining brand character and differentiation.

This certainly was a plausible explanation of our findings. The question now was whether to show the segmentation results in conjunction with the perceptual maps. After some discussion we decided not to show the segmentation results. We thought that they would be hard to interpret; essentially explaining the absence of differences is much harder than showing and explaining differences. In this case, it would be even harder since the argument was both long and subtle. And, since the results of the segmentation added little to our overall understanding of the appliance, not presenting them could be done with little loss. Parenthetically, I would argue against advancing very subtle explanations of data except when absolutely necessary. While we may appreciate our subtlety and cleverness in teasing out implications and formulating explanations, they can be lost on our audiences and can confuse them as well.

Assessing Advertising with Perceptual Mapping

Perceptual Mapping is often used to determine the actual or desired positioning of brands. The results of such analyses, as was the case in the appliance category, frequently become the basis for efforts at repositioning a brand in consumers' minds. We use perceptual mapping much less often to assess whether advertising is in fact positioning brands in the desired ways. This section will present results from a study which uses perceptual mapping to assess how advertising is positioning manufacturers.

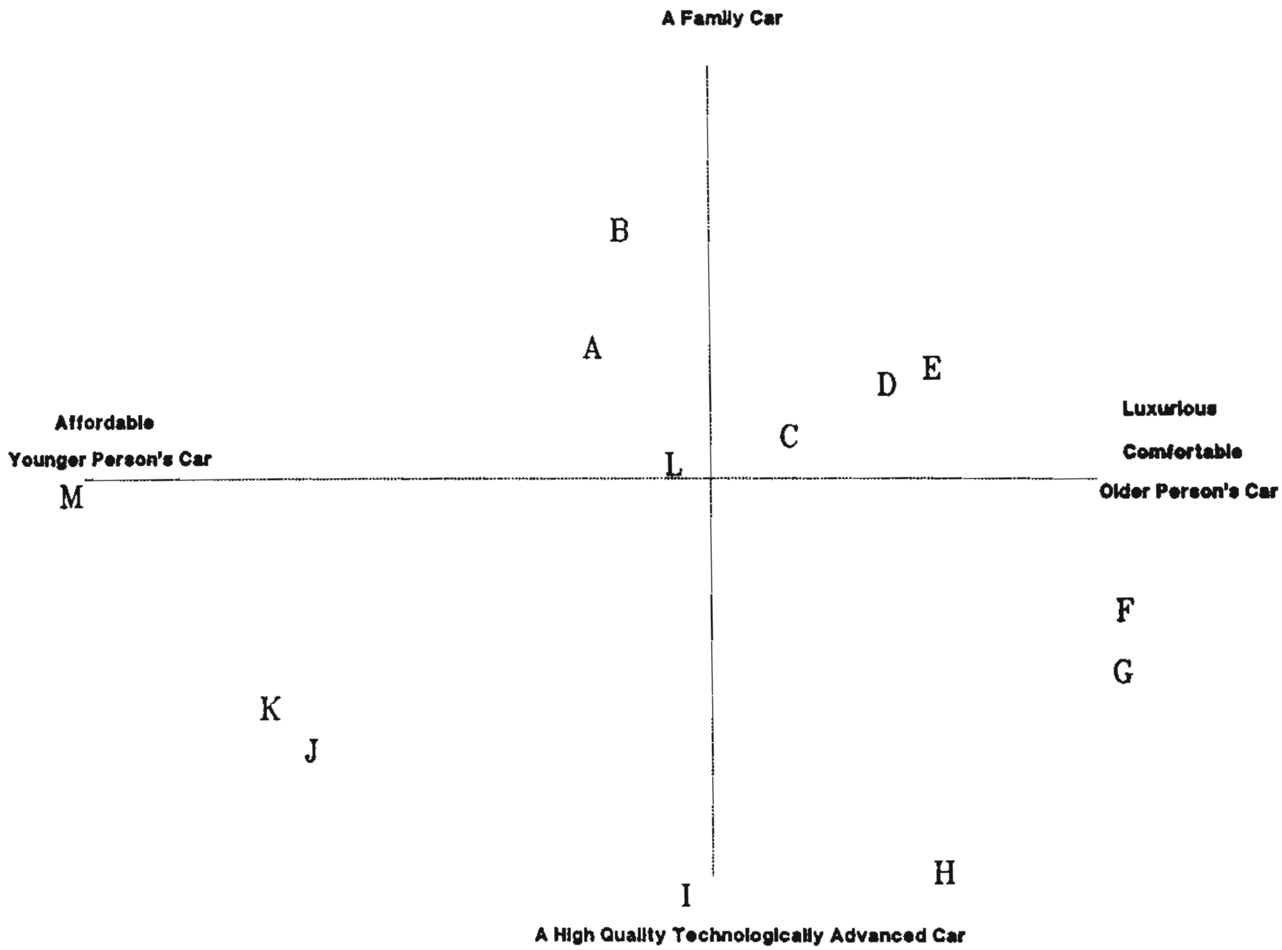
The data come from an ongoing study of the automotive category. The study is designed to assess the effect of advertising on the images or positionings of various manufacturers. The aim of the study is to determine in which direction on a map the advertising for specific manufacturers is moving the images of these manufacturers. Of course, not all directions are equal; the desire is that the advertising will be moving in a direction consonant with the desired and agreed upon positioning of the specific manufacturer, and this will be the only manufacturer moving in that direction.

In the study, respondents first rate several automobile manufacturers on 15 image attributes (quality, sporty, technologically advanced, and so on). They then see six commercials and read two print ads for several manufacturers and rate each manufacturer based on what the commercial or ad communicates about the manufacturer. The research is unique in that it attempts to assess simultaneously the effects of many campaigns, as opposed to individual commercials and ads, on the perceptions of many manufacturers.

There are two ways to determine the impact of advertising on the images or positions of automobile makers. Both ways begin with a map showing the structure of the market prior to exposure to the advertising. The structure is shown in Figure 1.

FIGURE 1

PRE-ADVERTISING MARKET STRUCTURE



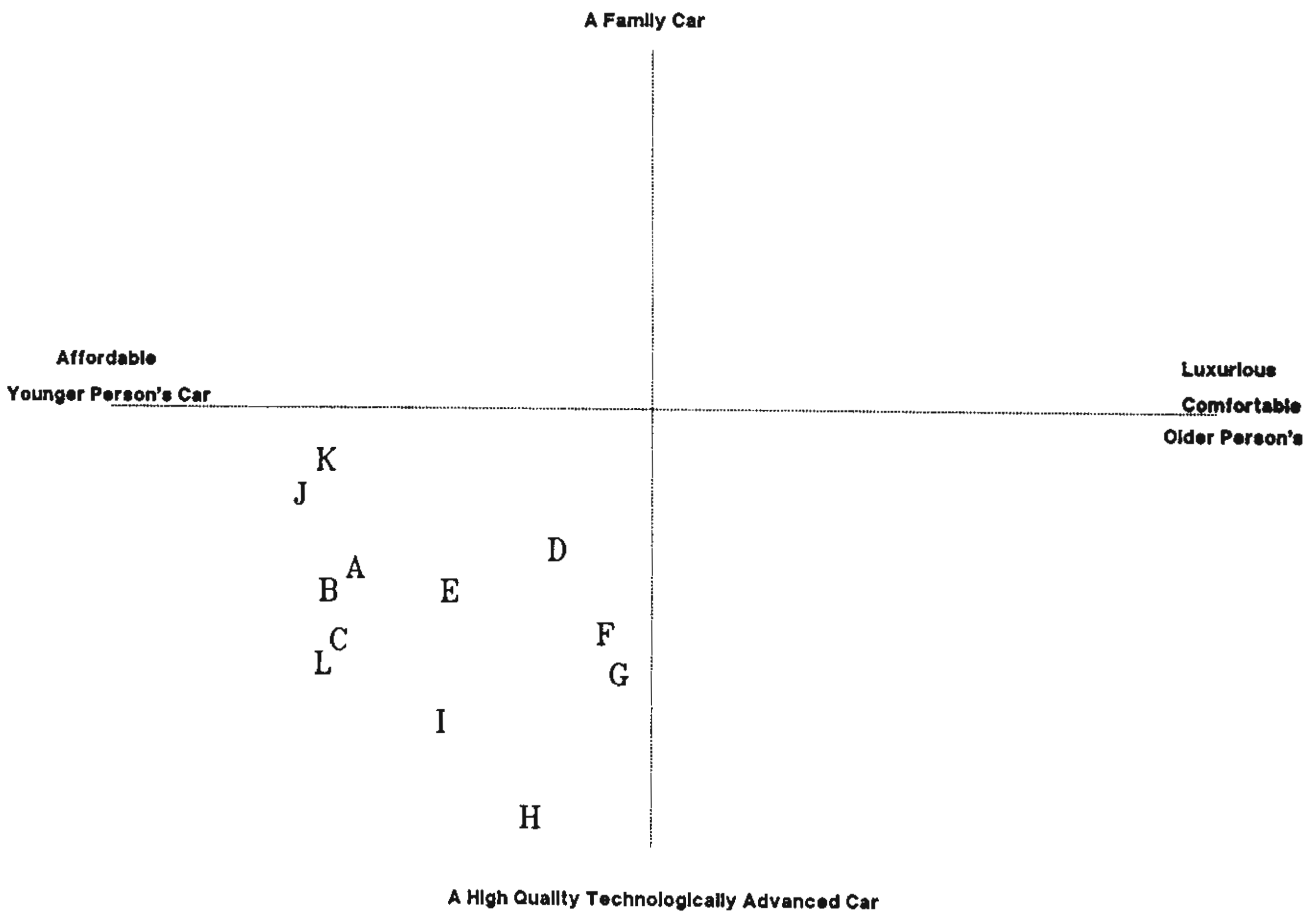
This map shows that people distinguish among manufacturers in the following ways. On one dimension they see cars that offer value and appeal to younger people; M, a low priced Oriental import best exemplifies this type of manufacturer. At the other end of this dimension they see cars that appeal to older people and offer more power and luxury; E, an American marque, is an example of such a maker. The other dimension has "technologically-advanced" and "high-quality" as its defining characteristics on one end, exemplified by a German marque, and family cars on the other end of the dimension, with an American marque being closest to this end of the dimension. Both the dimensions and the placement of the makers make sense to people familiar with the automotive category. From the map it appears as though consumers have fairly clear pictures of a number of cars. Where there is confusion in images, it is primarily among the American manufacturers who are the largest producers in the United States market and have had the greatest difficulty in differentiating the many models and brands they produce. The classification analysis bears this out. Overall we correctly classify 33 percent of the respondents, but the correct classification by maker varies from 13 percent for a domestic manufacturer to 74 percent for a foreign maker.

It is after exposure to the advertising that we have alternative ways of looking at and portraying the structure of the market. One option is to apply the original structure to the post-advertising ratings of each car; the other is to re-estimate the structure using only the post-advertising ratings. We have done this and the results of these two options are quite different.

Figure 2 presents the results of applying the original structure to the post-advertising ratings. This map is radically different from the market represented in Figure 1. All of the makes are now located in the lower left quadrant, whereas originally only makes I, J, K and O were there. Clearly, drastic changes have occurred, changes that most advertisers would not pleased with. Figure 2 implies much less differentiation among brands based on the advertising.

FIGURE 2

POST-ADVERTISING MARKET STRUCTURE

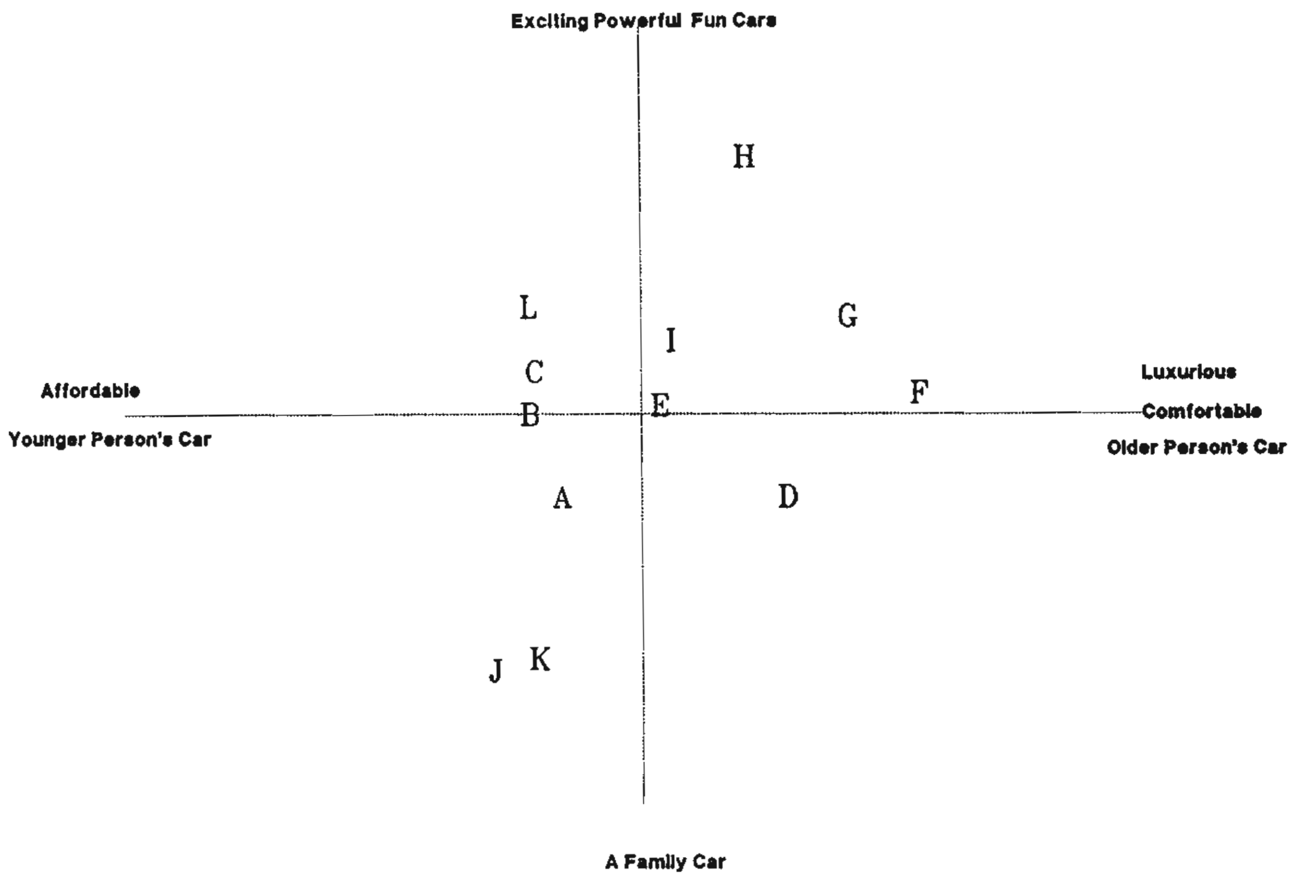


* ORIGINAL FUNCTIONS

This is apparent when we look at our ability to correctly classify people based on their post-advertising ratings of manufacturers. Whereas originally we could correctly classify 33 percent of the respondents, our ability drops to 6 percent based on the post-advertising ratings. Looking at the map it appears as though every maker's advertising is directed against the same strategy and communicating the same message.

FIGURE 3

POST-ADVERTISING MARKET STRUCTURE



* NEW FUNCTIONS

Figure 3 presents the results of re-estimating the structure using the post-advertising ratings. This produces a very different picture from Figure 2. There is greater dispersion among the manufacturers than in Figure 2. With the exception of maker I which in the original map was away from the center and closest to O, the general pattern seems similar to the original map. At least, we could all agree that this one might be based on the original structure, whereas we would be much harder pressed to agree with this contention statement regarding Figure 2. And our ability to correctly classify makers is not different from our ability in Figure 1, 34 percent versus 33 percent.

What do these two very different maps tell us? Should we use both to understand what is happening, or should we choose between them? I think each tells us something important about what the advertising is communicating about manufacturers and how it is working. Figure 2 tells us two things. First, it says that the original structure does not adequately describe the market based on the exposure to the advertising. There is virtually no differentiation among makers, and only 25 percent of the space is being used. At the same time, this map tells us something very important from a marketing sense. It says that everybody seems to be singing the same song about their cars. Everybody wants people to think their cars are youthful, offer good value for the money, and are technologically advanced and high quality. The net result is that manufacturers are blurring, rather than sharpening, their images.

Figure 3 tells us how the structure has changed based on the advertising, and thus provides insights into how the advertising is working. The horizontal dimension, the first dimension in this and the original solution, remains basically the same, describing characteristics of car that are seen to appeal to older and younger people. It is the second dimension that changes after advertising. Instead of being a family/affordable car versus high quality and technologically advanced car dimension, it changes to family/affordable car on one end to exciting, powerful quality car on the other end. This suggests that the advertising is attempting to change the relative importance of the criteria people use to judge cars. Another way of saying this is that the model of advertising as agenda setting appears to describe the way advertising is working in the automotive category.

By examining both maps, we have learned some important things about automotive advertising. There may be a lesson in this for mapping studies that are repeated at regular intervals. The lesson is that perceptual mapping can demonstrate the direction of change as well as changes in the underlying structure of consumers' perceptions of the marketplace. Each complements the other and adds to our understanding of consumers and the structure of the marketplace.

Summary

This paper discussed three different issues in positioning research. It offered a way of dealing with the potential problem of owners and nonowners producing spurious or misleading maps. The solution was to create owner and nonowner brands and estimate the space using owner and nonowner brands. We suggested that it was likely that one dimension would differentiate owners and nonowners and thereby eliminate their effects from the other dimensions. The paper also discussed using segments based on cluster analysis in maps. The example discussed showed that there may be occasions when it is better not to display the segments. Finally, the paper showed how maps can be used to assess advertising.

A Correspondence Analysis Approach to Perceptual Maps and Ideal Points

Vincent Shahim, Markinor, Randburg
Michael Greenacre, University of South Africa

1. INTRODUCTION

Correspondence analysis has been used by researchers in South Africa as a brand mapping technique for many years, but only fairly recently in the United States. This paper has two objectives. Firstly, we shall look critically at the various types of data used for brand mapping and the different mapping techniques which are available. Our general conclusion here will be that it is the type of data rather than the type of method which is the most important consideration. The other objective is to illustrate the versatility of correspondence analysis in its handling of alternative types of data, as well as its inherent ability to expose different aspects of a particular data set. We shall also touch on ways of extending the analysis beyond the basic brand maps that are commonly produced.

2. OBJECTIVES OF BRAND MAPPING

Why do we create brand maps at all? Some of the objectives of doing brand maps are:

- To display graphically the position of brands vis-a-vis other brands in the market.

- To identify the dimensions along which consumers perceive brands and then to display the brands along these dimensions. The dimensions that emerge will depend on the technique used and on the particular perspective in the data that we want to emphasize. Do we want the dimensions to focus on the importance of the attributes that influence brand choice? Perhaps they should reflect the attributes that differentiate the most between brands? Or should the dimensions reflect the correlations between attributes?

- To identify market gaps and optimal positionings for marketing strategy purposes. This could either be to position a new brand in the market or to reposition an existing brand.

- To identify the determinant or salient attributes in the market, i.e. the most important and the most differentiating attributes with respect to brand choice. This often takes us into the area of more formal statistical methods rather than a simple graphical display of data.

A graph is intended to summarize data and to display the data visually. A brand map used graphically is no different. Any summary results in a loss of a certain amount of information. It is hoped that the more important information is retained and the less important information is lost. When we do a brand map we should have a clear idea of the type of information we want to get out of the analysis. This should guide one in collecting the appropriate data and in choosing the appropriate method of analysis. Differing methods retain or emphasize different aspects of the data and the display itself gives only a partial view of the data. This can be likened to looking through a narrow window at a room full of interesting things (information). Depending on the situation of the window, one will get a different perspective of the room.

3. ATTRIBUTES

We should be clear as to what we mean by an attribute. Attributes can generally be divided into two broad categories. To be more precise we can place attributes on some sort of continuum. There are attributes that are very factual, such as product features. At the other extreme there are "attributes" that are much more in the area of pure image. Some of the ways that these two groupings have been labelled are

hard -- soft attributes
rational -- emotional attributes
tangible -- intangible attributes
attributes -- image statements

In some instances it makes sense to do separate perceptual maps for each grouping. On the other hand, it is often useful to see the interactions between the two groupings.

An attribute that differentiates among the brands is not necessarily important in influencing brand choice. On the other hand, an important attribute might not differentiate among brands because most of the brands are seen to satisfy that attribute. It has become a "generic" and is not determinant any longer. A determinant attribute is one that is important and differentiating. Sometimes an important attribute is not seen to be satisfied currently by any of the brands. It is thus not yet perceived to be differentiating the current brands. It is, however, potentially determinant. This is the clearest example of a "market gap."

Most mapping techniques, with multidimensional scaling being the exception, utilize brand attribute data. The map is therefore heavily dependent on the attributes used. It is important to identify the right attributes to use in order to produce a true picture of the market. Ideally, focus groups should be used to generate the attributes to be used in a quantitative study (in our experience, this is usually not done in practice).

4. CORRESPONDENCE ANALYSIS & OTHER TECHNIQUES

The majority of mapping techniques use the same mathematical tool to approximate a data matrix, namely the singular value decomposition (or SVD). What distinguishes these techniques from one another is the transformation of the original data prior to this approximation and the nature of the rescaling of the SVD solution to obtain the co-ordinates of the displayed points. Greenacre (1984, Appendix A) fully describes the unification of mapping techniques in the context of the SVD, including principal components analysis, correspondence analysis, discriminant analysis (or canonical variate analysis), canonical correlation analysis, and the biplot (see also Greenacre & Underhill 1982).

Discriminant analysis is a popular brand mapping technique in the U.S.A. and is usually applied to brand-attribute ratings. An advantage of discriminant analysis is that it attempts to take into account correlations between the attributes, but pays a penalty for this advantage by running into multicollinearity problems, as in multiple regression. A pre-analysis phase of either pruning the set of attributes or condensing them into a set of factors is necessary, the latter option being a popular approach in American applications.

Correspondence analysis distinguishes itself from all the other techniques in this class by being oriented toward categorical data rather than measurement data, although it can also be used to analyze measurement data in a suitably coded form. In fact it is the only technique of its kind which can handle nominal categorical data, i.e. categorical data where there is no inherent ordering of the categories. Since both brand-attribute associations and ratings are categorical in nature, correspondence analysis can handle both naturally. In fact, we see the versatility of this method as its strongest point. Its ability to be a jack-of-all-trades means that the researcher can standardize on one methodology to solve all his mapping problems. This is the philosophy of the French school of data analysts who were the originators of the method and who encourage the user to transform and reweight the data to focus on different features, rather than apply different methods to the (untransformed) data.

Important features of correspondence analysis are the reweighting concept and the method's ability to display additional information in the form of so-called supplementary points. These concepts, however, are not special to correspondence analysis and can be implemented in any of the mapping techniques discussed here. It is in the framework of correspondence analysis, however, that reweighting and supplementary points have been the most extensively and fruitfully used. (This is analogous to factor analysis where rotations are standard practice to improve the interpretability of solutions - this is not peculiar to factor analysis and a correspondence analysis solution can just as easily be rotated if so required.)

Another aspect of correspondence analysis of importance here is whether to perform a so-called doubling of the attribute data or not. Doubling of ratings is considered standard in correspondence analysis (see Greenacre 1984, Chapter 6: Analysis of Ratings and Preferences), and consists of creating a pair of points for each attribute, in effect its "positive" and "negative" poles. The decision whether to double depends on whether we wish to expose an effect of overall level of attribute association or rating, also called the "size" effect. The size effect is inextricably bound up with the so-called "halo effect" and usually appears as the first dimension in an unrotated principal components analysis or factor analysis. An ordinary correspondence analysis partials out the size effect in a certain way and doubling is an attempt to reintroduce it in the analysis.

Finally, as a contrast to the above techniques, multidimensional scaling does not require brand-attribute data to produce a map. The map is based on respondents' perceptions of brand similarity. The data thus consist of some type of similarity score between each pair of brands. The chief characteristic of this approach is at the same time its greatest advantage and greatest disadvantage. The advantage is that the dimensions along which brands are plotted are not based on attributes defined by the researcher, but are implicitly respondent defined, based on their perceived similarity between the brands. The disadvantage is that we do not know what these dimensions are -- it is left to the researcher to name the dimensions, either based on his or her knowledge of the market or on other data collected in the survey.

5. RETAIL CLOTHING STORE SURVEY

To illustrate the application of correspondence analysis, we conducted a survey among women in the Johannesburg area of South Africa to establish the usage and perception of the major retail clothing chains. The population was defined as women who have purchased clothing for themselves from any of seven selected chains in the last two years. A split sample was used, with each sample having matched respondents and responding to a different questionnaire. Interviews were conducted in four selected shopping centers. The method of interviewing used was a computer-aided personal interview. Final samples of 101 and 102 were obtained, giving a total of 203.

The attributes on which the stores were surveyed are summarized in the following table.

TABLE 1

Twenty attributes for clothing store survey.
The letter codes are used in the graphs and brand maps shown later.

A	– sells fashionable clothes	K	– reasonable prices
B	– offers value for money	L	– wide range of clothing
C	– caters for sophisticated people	M	– attractive store design and decor
D	– friendly staff	N	– efficient service
E	– warm appealing atmosphere	O	– attractive presentation of merchandise
F	– low price orientated	P	– easy to find what you are looking for
G	– caters for all shapes and sizes of people	Q	– is becoming more popular
H	– sells trendy clothes	R	– sells high quality clothing
I	– caters for older people	S	– sells clothes for working women
J	– caters for younger people	T	– is improving and moving ahead
	–	U	– is your type of store (ideal)

5.1 Questionnaire

The questions that were common to both samples were:

- Stores used in last two years.
- Stores used most often to fourth most often.
- A familiar store which would not be considered in the future.
- Preference ranking of stores.
- Constant sum paired comparison on preference with up to four stores being compared based on usage and "not consider." The brands that enter this exercise for each person are called the evoked set.
- Importance rating of attributes.
- Importance ranking of attributes from first to fifth most important.

Questions peculiar to the first sample (or the "Association Sample") were:

- Brand-attribute associations on the 20 attributes plus associations of the ideal store.
- Attributes also considered important for ideal store in addition to those ranked first to fifth.

Questions peculiar to the second sample (or the "Ratings Sample") were:

- Brand-attribute ratings on a seven-point scale on the 20 attributes for all stores included in the evoked set, plus a rating of the ideal store.

The most important difference between the two samples was the way in which the brand-attribute data were collected.

5.2 Objectives of the survey

The main objective of the survey was to contrast the brand maps produced using different types of data and data transformations. In addition, we wanted to examine different methods of producing ideal points on maps and of identifying the most important and the determinant attributes.

Some of the questions that we wished to address were:

- Are brand-attribute associations sufficient or do we need to do brand-attribute ratings?

- Do different data codings produce markedly different maps? In particular, is it sufficient to use summarized data for data input or do we need to input individual respondent data?

- Does the identification of ideal regions of the map depend on the data that are used to locate the ideal? In particular, how do the ideal points produced using importance ratings, importance rankings, preference ranking and paired comparison preferences differ?

- On what data should the map be based, on the brand-attribute data, the importance data, or is there some alternative way of doing it?

The stores will be referred to throughout this paper as Stores 1 through 7.

5.3 Data

The data that we have available for mapping constitute a three-way matrix of respondents by brands by attributes. To analyze this data matrix by existing methods, we shall have to collapse two of the "ways" into one. We adopt the most common approach by collapsing respondents and brands to define the rows of the two-way data matrix, with the attributes defining the columns. The structure of this data matrix is depicted in Figure 1, while the condensed version of the raw data into a frequency table is depicted in Figure 2.

Looking at these matrices one can see that there are a number of approaches that can be used to produce a brand map. The map can be based either on the individual respondent data (Figure 1) or on the condensed data (Figure 2). While it is preferable that the individual respondent data be analyzed, we shall also analyze the condensed data for purposes of illustration.

We can see four possible ways to generate a brand map:

1. Use the brand-attribute data to generate the basic map, with the importances, ideals, and preferences superimposed on the map afterwards.

2. Use the attribute importances to generate the basic map, with the brands being superimposed afterwards.

3. Use the brand-attribute associations (or ratings) and the attribute importances jointly somehow to define the map.

4. Use the brand-attribute data and the brand preferences jointly to define the map.

Each of these four ways focuses on a different aspect of the data. Here we restrict ourselves to a discussion of 1. and 3. only as illustrations of our approach.

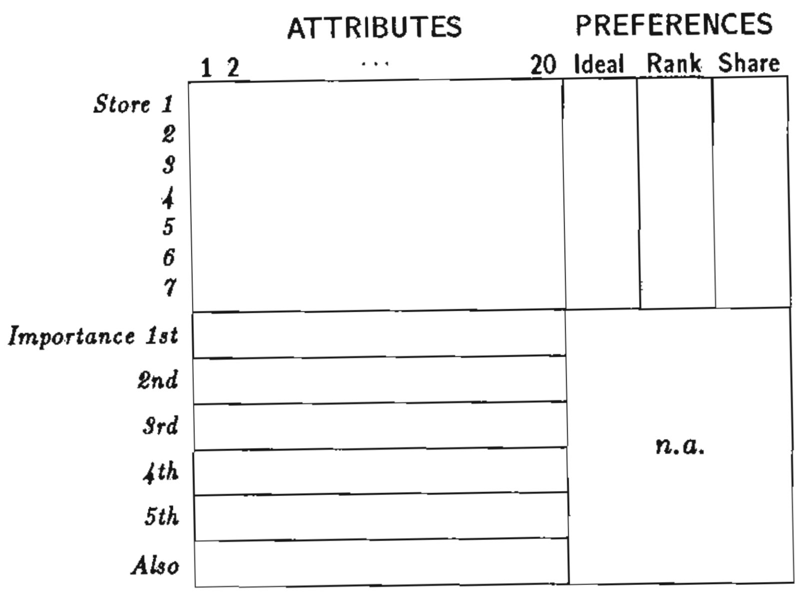
FIGURE 1

DATA MATRIX IN RETAIL STORE SURVEY

		ATTRIBUTES	PREFERENCES		
		1 2 ... 20	Ideal	Rank	Share
Respondent 1	*Store 1*				
	Store 2				
	Store 3				
	Store 4				
	Store 5				
	Store 6				
	Store 7				
	Importance		*n.a.*	*n.a.*	*n.a.*
Respondent 2	*Store 1*				
	Store 2				
	etc...				

FIGURE 2

CONDENSED MATRIX OF FREQUENCIES IN RETAIL STORE SURVEY



6. EXAMPLES OF CORRESPONDENCE ANALYSIS

In this section we show some of the maps we obtained by analyzing the store survey data in different ways by correspondence analysis. All these analyses were executed using the microcomputer program SimCA 1.5 by Greenacre (1986).

6.1 Plain vanilla version

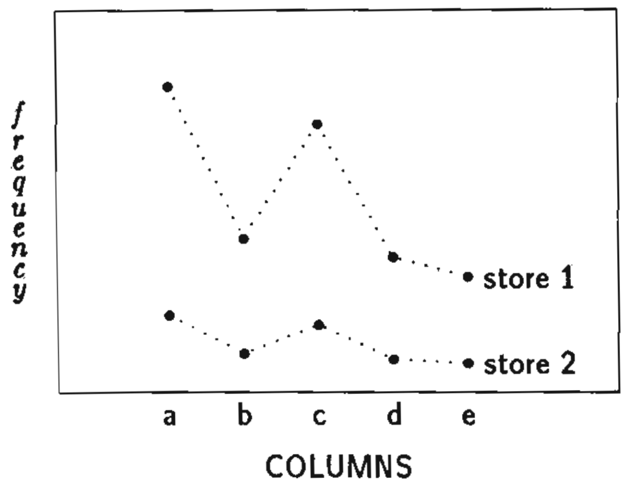
In South Africa, the data often used to create a correspondence analysis brand map are the summarized brand-attribute associations. The brands are the rows of this matrix and the attributes the columns. Each cell of the matrix contains a count of the people who associated the brand with the attribute.

Figure 5 shows the map obtained by analyzing the data in the form depicted in Figure 2. To encourage the correct interpretation and to make the displays less complex we have separated the rows and the columns into two maps (these are usually overlaid as one map). With reference to Figure 2, it is important to bear in mind that the row points of the matrix, i.e., the brand positions and the importances, are in the same space (Figure 5a) and the distances between any of these row points on the map can be compared. In the same way the attribute points, the ideal brand point, and the preference points are in the same space (Figure 5b). We stress this because some recent articles by Carroll, Green & Schaffer

(1986, 1987) have suggested that with a different scaling of the dimensions it is possible to interpret the distances between all points (both row-to-row, column-to-column and row-to-column). This approach has serious deficiencies, however, as pointed out by Greenacre (1989).

FIGURE 3

TWO ROWS WITH SIMILAR PROFILES ACROSS COLUMNS



Each brand point is determined by its profile across all attributes, i.e. in this case its set of frequencies for the attributes divided by the total of these frequencies. An important point to note is that the absolute level of associations of a brand is not reflected in this display. As illustrated in Figure 3, two brands with similar profiles - one at generally high levels and the other at generally low levels - will have similar positions on the map. The aspect of the data that the display emphasizes is the relative positions of the brands or each brand's relative strengths and weaknesses. The overall "goodness" or "badness" of each brand is not displayed. We say that there is no size dimension displayed, only dimensions of shape.

The interpretation of the display in Figure 5 is performed from the center outward. An attribute close to the center is not differentiating the brands, and the further one moves from the center outward, the more differentiating the attribute. Brands are interpreted in the same way. For more details on the interpretation of correspondence analysis see Greenacre & Hastie (1987).

The "halo effect" is commonly observed by researchers. A brand that has a general positive image will often score well on all the attributes, and vice versa for a weak brand. This method of mapping removes the halo effect in a particular way and thus allows us to concentrate on the brand differences. Notice that this is not the only way of "partialling out" the halo effect, neither is it necessarily the most satisfactory way. In removing the halo effect we do not really know how much other relevant information we are also losing. In fact, it is often relevant to know what the overall level of association or image a brand has. In the subsequent sections we discuss approaches to mapping that do not hide the halo effect.

6.2 Ideal points

The capability of correspondence analysis to plot supplementary points has been used to superimpose some of the other data on the map, to identify ideal regions and important attributes. This is done using two basic methods, the inferred importance method and the self-explicated importance method.

The preference ranking of brands runs clearly from the top left to bottom right of the display (Figure 5b). We infer that attributes in this direction are the more important ones in influencing choice (inferred importance method). The ideal attribute Is your type of store (U on the map) is also in this direction. The three measures of attribute importance are also plotted (Figure 5a). Average importance rating (Imp+), five most important taken together (Imp5), and five most important plus also important (ImpA), are all in the same direction. This would indicate that any of these measures can be used (self-explicated importance).

We notice that the ideal region based on attribute importance is in a slightly different direction than the one based on brand preference. It should be remembered that store preference gives an ideal that is based on existing brands. But none of the existing brands might actually be satisfying respondents' perceptions of an hypothetical ideal. This method indicates preferred regions in terms of current brands but does not really find gaps in terms of attributes not currently satisfied.

The importance data identifies ideal regions in terms of what respondents say is important. We notice that the importance points are being pulled in the direction of the price/value area of the map (attributes F K and B) as well as toward the bottom right quadrant. People tend to become too rational, particularly with regard to price. What people say is important when we directly elicit importance is not always the same as importance that is inferred from other data. We refer to the former method as <u>self-explicated importance</u> and the latter as <u>inferred importance</u>. Both approaches to identifying ideals provide valuable information, but each has its drawbacks.

The most important criticism of this approach is again the use of condensed data. This problem is the most acute when we think of how the ideal points based on the importances and preferences are being plotted, because these have been condensed and analyzed quite separately from the brand-attribute data. If a high proportion of people (say 50%) indicate that brand X is their most preferred brand, and (separately) 50% of the people say that brand X has a certain attribute, then it does not necessarily mean that these two sets of people are the same. In fact, theoretically they might not even overlap! The mapping of the ideal preference point toward this attribute could be totally erroneous. We should really be examining the relationships at the individual level between brand-attributes and importances or preferences.

6.3 Doubling

The map in Figure 6 illustrates a simple method of introducing an axis of size onto the map. This is accomplished by doubling the matrix of associations to produce a positive and negative point for each attribute. To do this one subtracts the complement of the sample from each association. This clearly shows an axis of size along the horizontal dimension, with the brands lining up from good to bad on most attributes (from left to right) along this axis. The vertical dimension is mostly high price (F-, K-) versus low price (F, K). In this particular example we lose only a little of the shape information when we introduce a size dimension, since the positions of the attributes from left to right in Figure 5 appear to be similar to the positions of the positive points from top left to bottom right in Figure 6. It is more common to find that the display with respect to axes 2 and 3 of the doubled analysis is similar to the one with respect to axes 1 and 2 of the undoubled analysis.

The only criticisms of this data doubling are that we are artificially creating the negatives, and are implicitly assuming that if respondents do not associate an attribute with a brand, they would associate the brand with the negative of that attribute. In fact, the so-called "negatives" here are really "non-associations" or "non-mentions."

6.4 Ratings

We now repeat the analysis using the ratings data. Again we are only using the summarized data, this time a matrix of average ratings of brands by attributes. In this situation it is perfectly acceptable to double the matrix -- it is in fact theoretically more correct to do so. The respondent rates the brands from positive to negative on each attribute (on a seven point scale in this case). Again the first axis is a very clear size axis (see Figure 7). The difference in the configuration in this display and the one using the doubled association data can be attributed to the large number of "not consider" respondents that rated Store 3. If we in fact remove this column from the analysis we end up with a very similar display to the one using the associations.

In this case, and in fact in the other analyses we performed comparing the association and ratings data, the association data produces a display that is at least as good as the one produced using the ratings data. Using the association data also avoids the distortion problems that the ratings data could produce depending on which subsets of brands we get each respondent to rate (cf the Store 3 factor just described).

6.5 Individual data and confidence intervals

We now turn our attention to using the data at the individual respondent level. Considering the association data first, this involves a zero/one coding of each brand on each attribute for each respondent. While an undoubled matrix can be used, we have concentrated our analysis on the doubled data. The resultant configuration of the attribute points in this example (Figure 8) is very similar to that produced using the summarized data (Figure 6). Although it is theoretically preferable to use the individual-level data rather than the condensed data, it is our experience in four different applications that the two approaches have lead to only slight differences in the attribute positions. This will not necessarily be the case in general, though, and we advise using the more detailed data even though this is computationally more intensive. The microcomputer program by Greenacre (1986) has been updated recently, with the specific intention of handling thousands of cases so that individual-level data can be directly analyzed.

Because we are analyzing the raw data, there is actually a position for each brand for every respondent (i.e. 102 x 7 = 714 brand points). It is sometimes informative to plot these so that one can see the groupings and overall scatter of brand points; the usual approach is to plot only the positions of the average brand points, resulting in only one point for each brand.

However, we can introduce a further refinement, thanks to the data being mapped at the individual respondent level. It is possible to derive a confidence region for each brand. This is approximated by an ellipse which shows the region where we can expect the average position of the brand point to fall with high confidence, as shown in Figure 8a. We notice that the ellipses for brands Store 5 and Store 6 overlap which means that one cannot be sure that their positions are different. These ellipses will obviously become smaller as the sample size increases. These confidence ellipses solve one of the problems of brand mapping by displaying the statistical difference between brands.

6.6 Weighted Spaces

So far in these examples we have been using brand-attribute data to produce the maps. The display thus focuses on the brand differences in terms of attributes and not on the importance of attributes. While we can plot the importance data as an ideal point as shown previously, these data do not influence the configuration of brand and attribute points. It is merely superimposed or fitted to the given display as best as possible. We could also have based the map on the importance data and then fitted the brands onto the dimensions created by importance.

However, remembering our definition of a determinant attribute as one that is differentiating and important, one might ask the question whether it is not possible to get our map to display this information. Our experimentation in this area has resulted in a possible way of doing this using correspondence analysis.

This is accomplished by weighting the brand-attribute data for the respondents by their respective importance data. If an attribute is associated with a brand, then that association is weighted proportionally to the importance the particular individual attaches to that attribute. This weighting can be done using the importance rating or the importance ranking of attributes. On the one hand the importance ranking forces the respondent to make a choice among attributes, with the result that fewer attributes tend to emerge as important. On the other hand, the ratings tend to result in quite a few of the attributes emerging as important.

The map in Figure 9 was produced using the association data reweighted by the importance ranks. Comparing this map to the one in Figure 8, we see that the result is precisely what we were hoping for. The highest-ranked attributes move outward away from the center while the less important attributes move toward the center (cf Figure 4 to see average importances of attributes). The dimensions now reflect attributes that are important and that differentiate.

We notice that the ideal points PrefRank, PrefShare, and attribute U ('your type of store') are all in the same direction. From this, we infer that attributes in the lower right quadrant are the more determinant. This way of relating the preferences to the dimensions is very similar to the preference regression techniques often used after the fact to establish the importance of brand map dimensions.

One worrying feature of this map is that the attribute B ('good value') tends to dominate the display because its importance ranking is much greater than any of the other attributes. This is relevant information, but tends to hide some information on other attributes because of its dominance. The solution to this problem illustrates a useful strategy that good "mappers" often adopt, simply to drop this attribute and repeat the analysis.

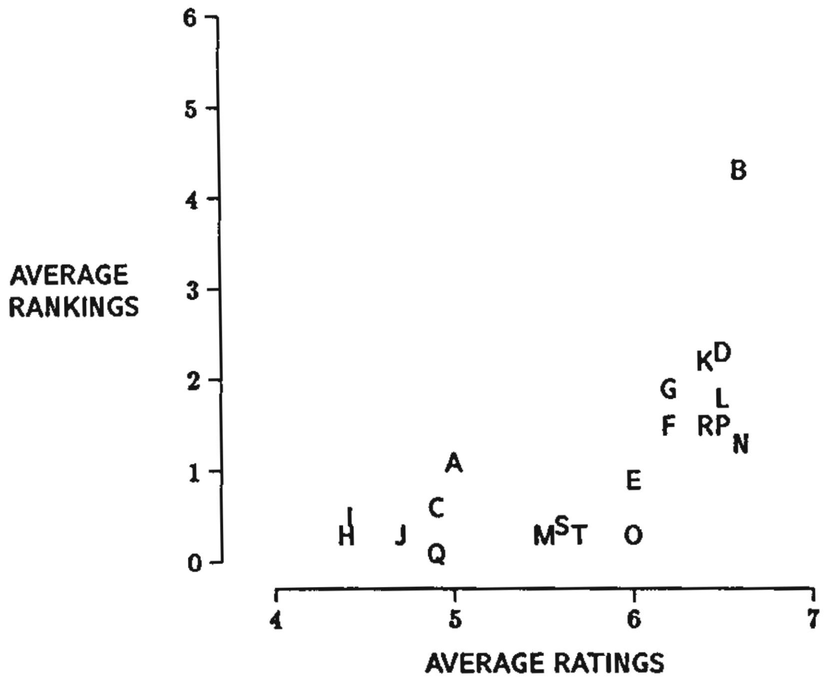
We can also project ideal points onto any of these maps, using either the preference data or the importance data. Each individual respondent now has an ideal point which can be plotted. This is more useful than the average ideal point produced when using the summarized data. After examining the scatter of ideal points, we can either find the average ideal point or produce average ideal points for segments of respondents. Figure 9 also includes examples of individual ideal points.

The analysis can also be performed using the importance ratings as weights. Although not reported here, the result is that fewer attributes move toward the center than in the more extreme example using the ranks as weights. This is because more attributes have high importance scores.

In exactly the same way, we analyzed the individual brand-attribute ratings weighted by the respective importance rankings (or ratings). Again we found that the analyses of the ratings did not yield any additional information on the relative positions of brands, attributes, and ideal points.

FIGURE 4

IMPORTANCE OF ATTRIBUTES



ABBREVIATION KEY TO FIGURES

The following abbreviations are used for the Figures that follow:

A – fashionable	H – trendy	O – attractive presentation
B – value	I – older people	P – clothes easy to find
C – sophisticated	J – younger people	Q – becoming more popular
D – friendly staff	K – reasonable prices	R – high quality
E – warm atmosphere	L – wide range	S – working woman
F – low prices	M – attractive decor	T – improving
G – all shapes & sizes	N – efficient service	U – your type (ideal)

Pref1...5 – 1st...5th brand preferences	*Imp+* – average importance rating
PrefRank – brand preference ranks	*Imp5* – 1st to 5th most important attribute
PrefShare – brand preference share – (from paired comparisons)	*ImpA* – 1st to 5th most important plus also – important

FIGURE 5

CORRESPONDENCE ANALYSIS OF STORE–ATTRIBUTE ASSOCIATIONS

Analysis of frequency table, undoubled

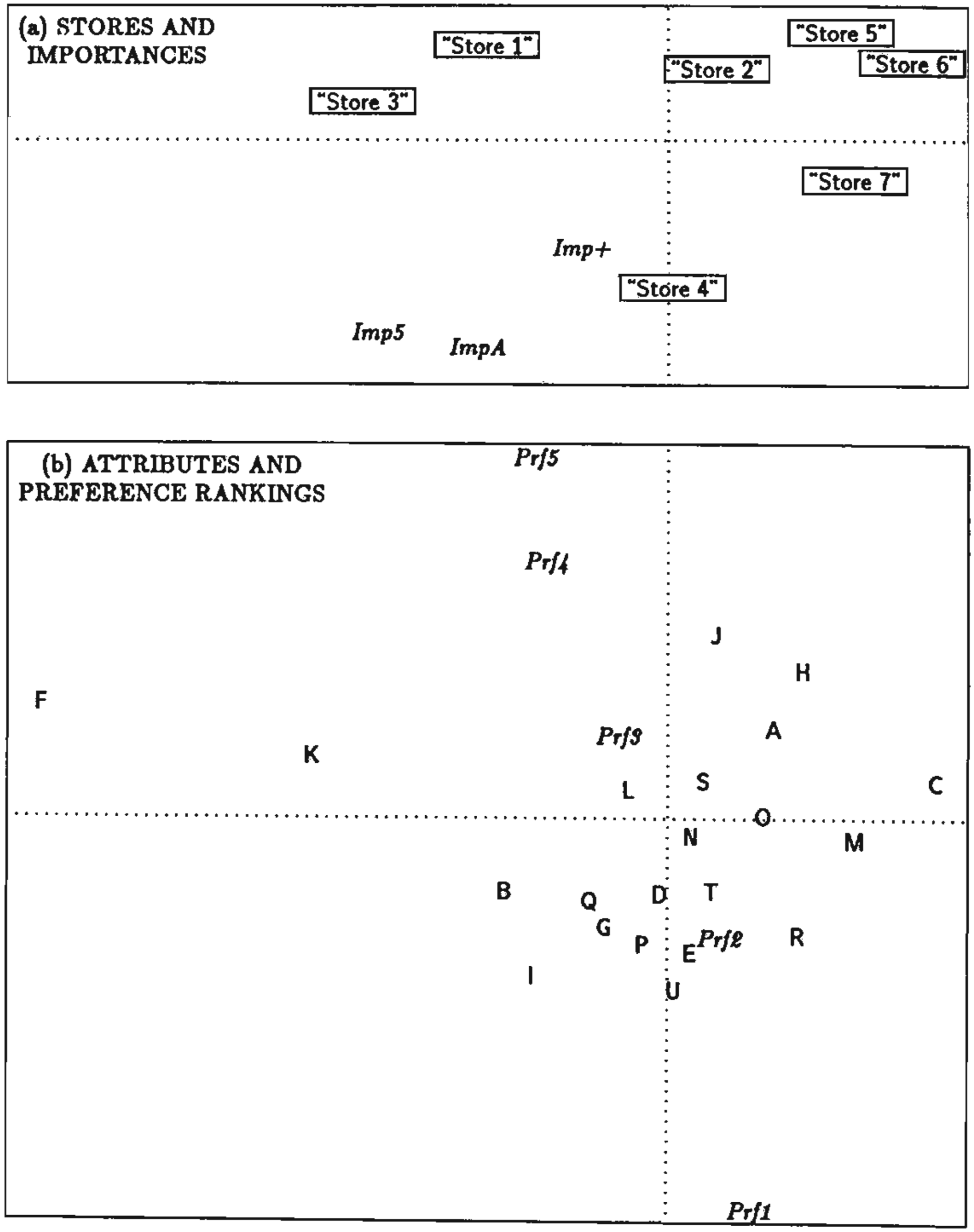


FIGURE 6

CORRESPONDENCE ANALYSIS OF STORE–ATTRIBUTE ASSOCIATIONS

Analysis of frequency table, doubled

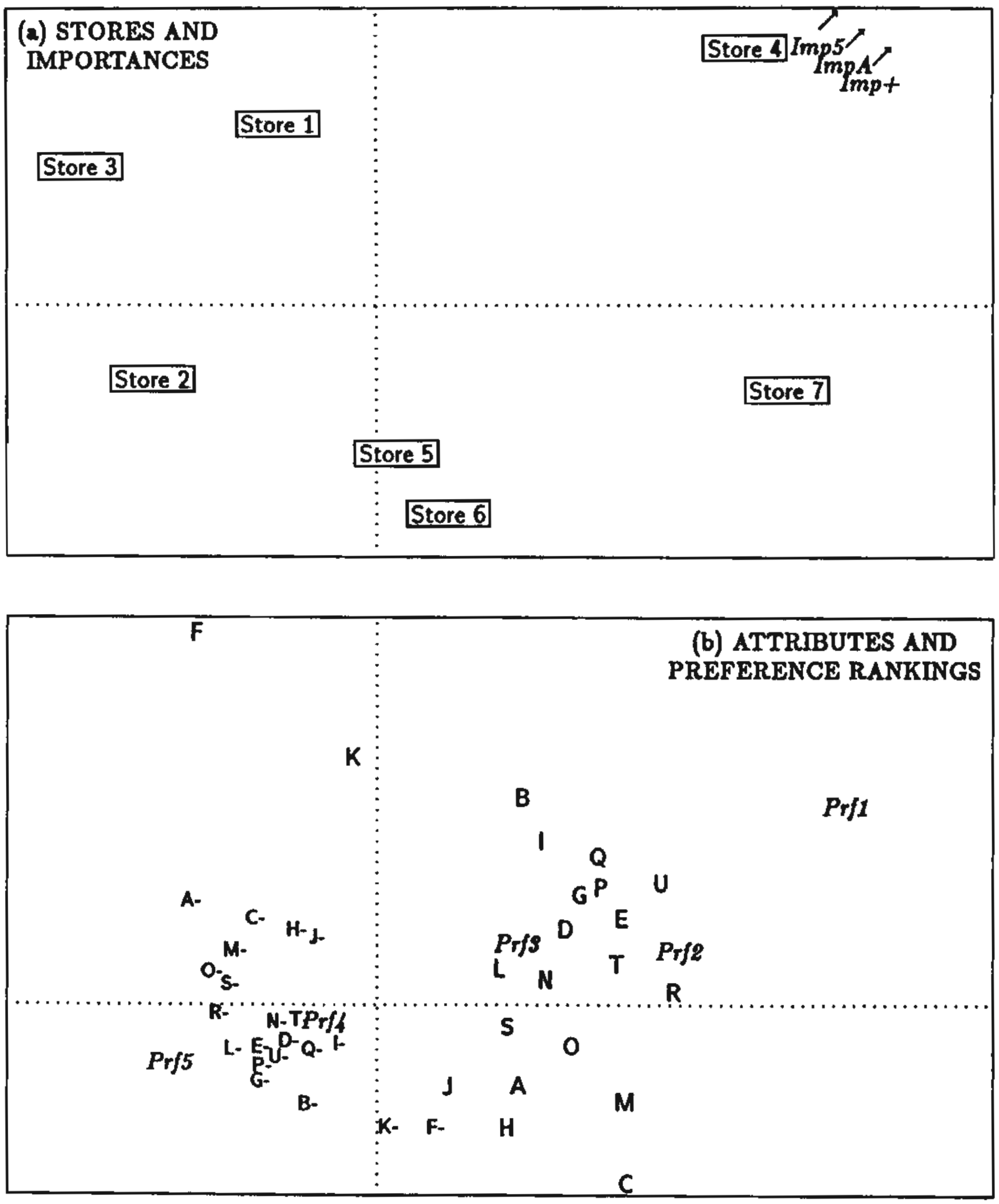


FIGURE 7

CORRESPONDENCE ANALYSIS OF STORE–ATTRIBUTE RATINGS

Analysis of condensed data, doubled

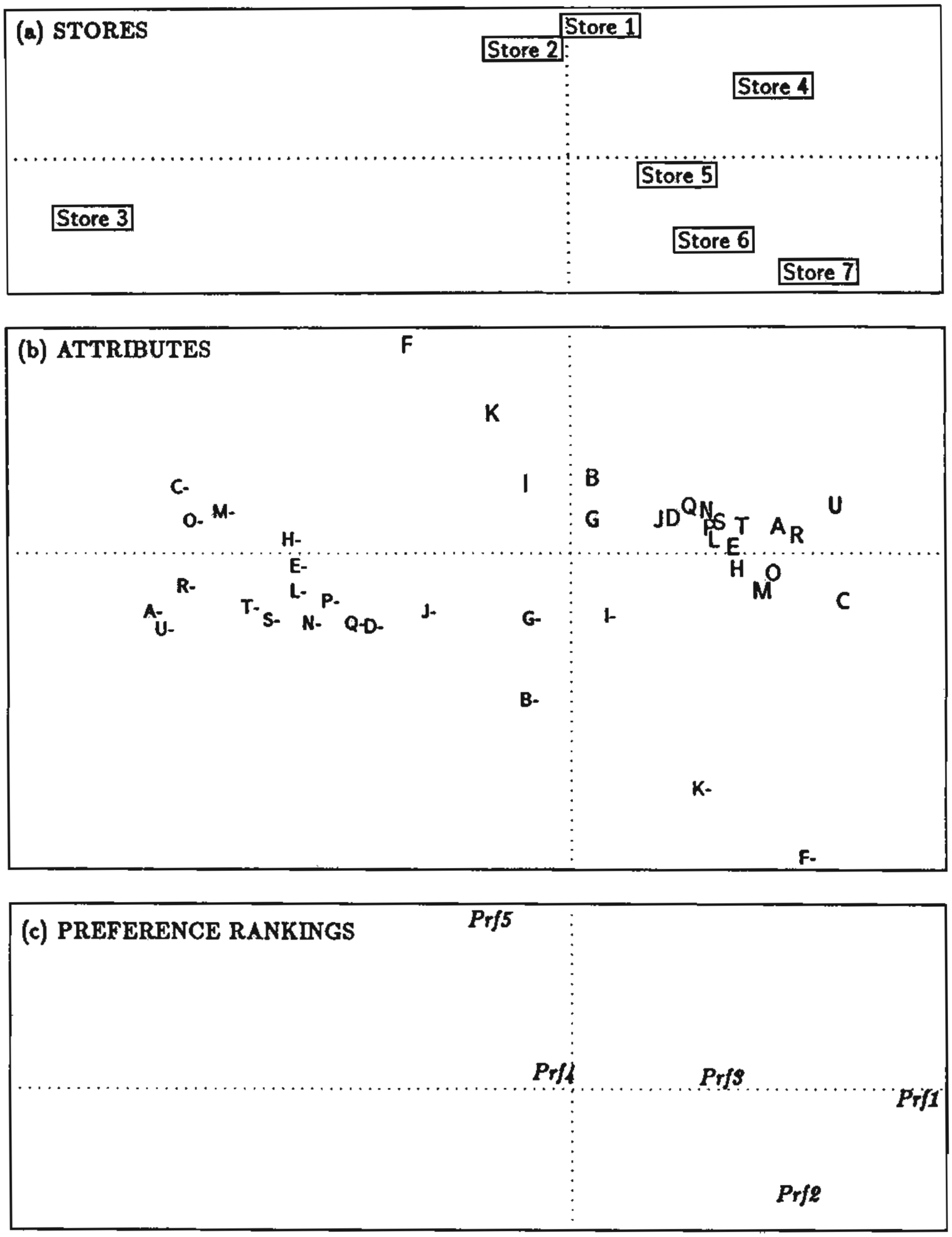


FIGURE 8

CORRESPONDENCE ANALYSIS OF STORE–ATTRIBUTE ASSOCIATIONS

Analysis of individual respondent data, doubled

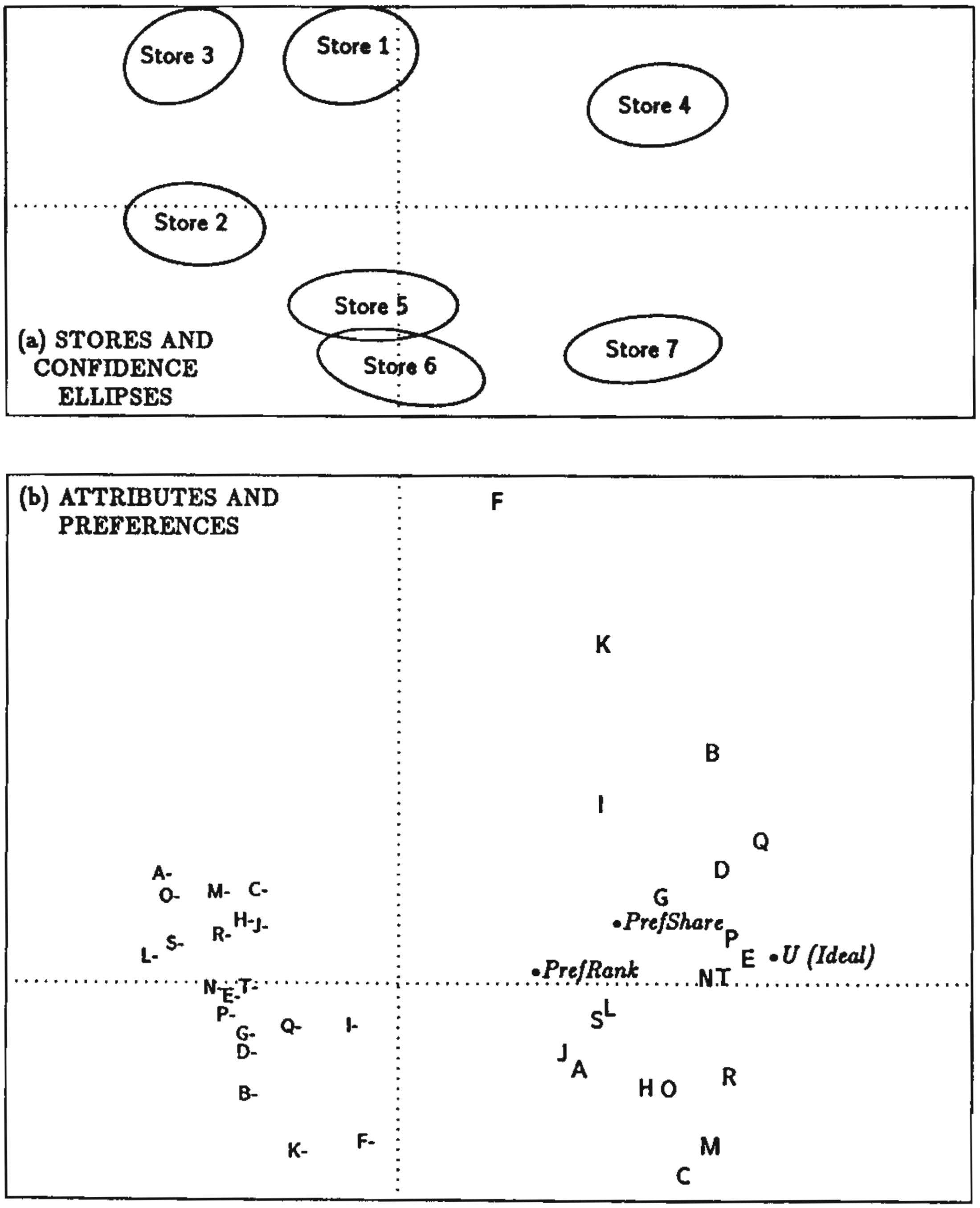
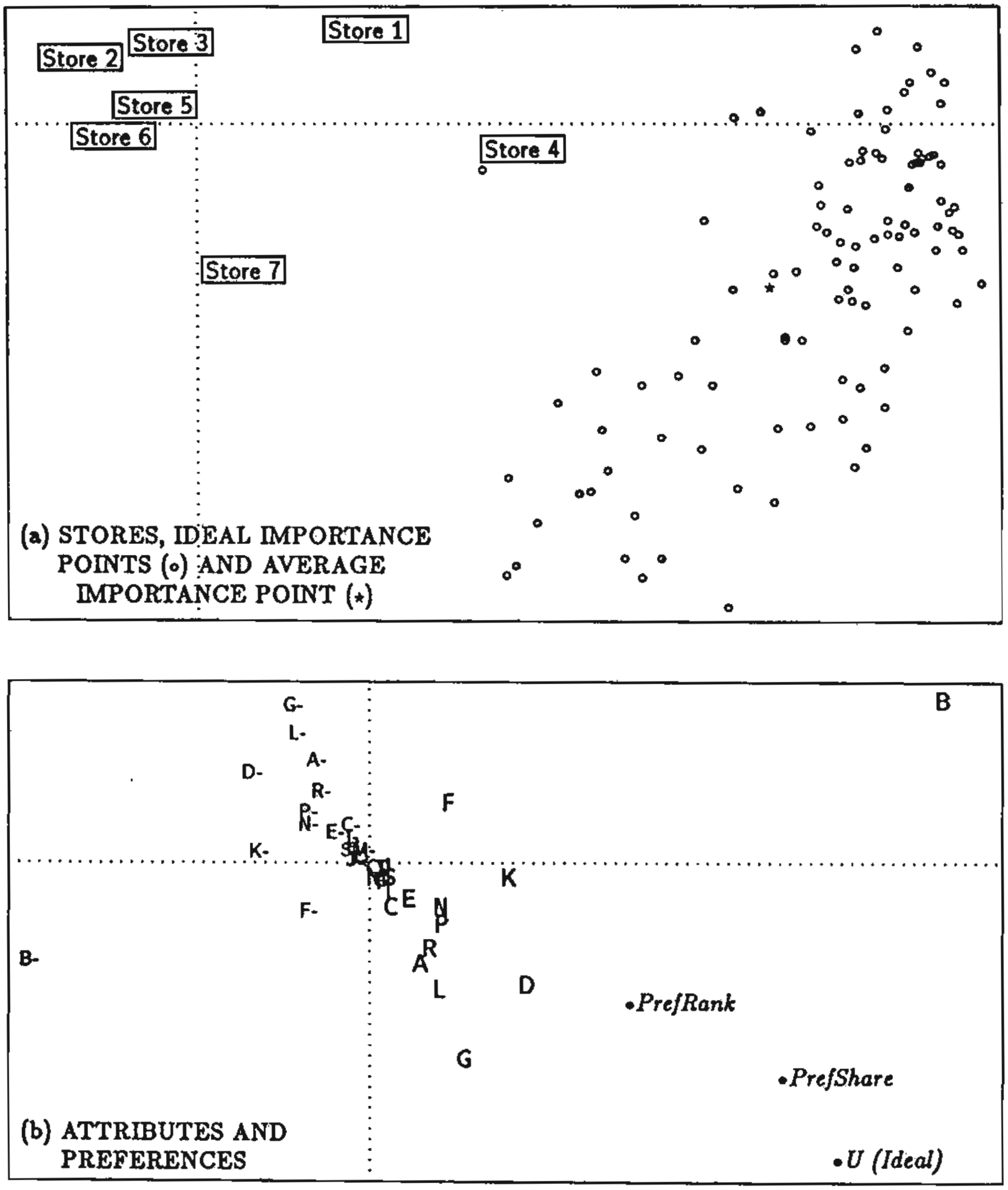


FIGURE 9

CORRESPONDENCE ANALYSIS OF STORE-ATTRIBUTE ASSOCIATIONS WEIGHTED BY ATTRIBUTE IMPORTANCE RANKINGS

Analysis of individual respondent data, doubled



7. DISCUSSION

7.1 Negative attributes

As mentioned previously, doubling brand association data might suggest that a brand not associated with an attribute is perceived to be negative on that attribute. This is not necessarily so and the "negative" pole created by the doubling is really a "non-association" category. A respondent will often not mention a brand because he does not know much about that brand, not because he has a negative image of it. The result is that the low awareness brands end up on the negative side of the size axis if we double the associations. The correct interpretation of brands at this end of the axis, however, should not be that they have a negative image, but rather that they receive few associations on the attributes.

A simple solution to this problem is to ask for a positive and negative association on each attribute. For example, "Which of these are high quality brands?" and "Which of these are not high quality brands?" This, in effect, creates three categories for each brand on each association, "good," "bad," and "neutral/don't know." The respondent's task is now longer but is still considerably shorter and not as tiresome as asking for a rating of a number of brands on each attribute. This is intuitively a good method which has proved successful in a number of applications. Using this type of data collection separates the low awareness brands from the negatively perceived brands, which is impossible to do if just positive associations are elicited.

7.2 What is the best way to map?

Our experience is that there is no one best method of brand mapping. Different methods (in particular different data transformations or weightings rather than different statistical techniques) give one different perspectives of the data. All of the methods can provide useful information.

Our preference is, however, for brand associations rather than brand ratings. The maps produced are often very similar, with associations resulting in maps that are at least as good and perhaps better than those using ratings, particularly if we are not able to get every respondent to rate all the brands. This will depend on the number of brands (or relevant brands) in the market. Researchers are well aware of the degradation in the data resulting from asking respondents to rate too many brands. If we ask each respondent to rate a subset of the brands, the problem is that the way this subset is defined can radically alter the data and thus the map. If we limit the brands to the most preferred ones, we miss out on those that are negatively perceived. Including negatively perceived brands (brands not considered) could also cause distortions, as happened with Store 3, and also pushes up the number of brands each respondent must rate.

In addition, while we feel that further experimentation is required, our preference is to use importance rankings rather than ratings. A good method seems to be to ask for a ranking of the first five most important attributes and then ask which other attributes are also important. This forces a choice between attributes and is also a slightly shorter task for the respondent.

Finally, the method illustrated of producing importance weighted spaces shows a lot of promise, but needs to be tested on a lot more surveys in practice. An important consideration is also whether the results are understandable and can be effectively communicated to users of research.

7.3 Determinant attributes

The importance-weighted space and the ideal point directions have given us a good indication of the determinant attributes (see Section 6.6 and Figure 8).

The thread that runs through the all the analyses is that there are certain attributes which emerge from direct elicitation of importance and that these are not always the same as attributes whose importance is inferred from other data (e.g., preference). The first type are often the more rational type attributes (e.g., good value), while the latter are often more "image" or emotional type attributes.

In our view, this is not a major problem as long as we are able to identify the two types. In terms of the theories of consumer behavior a consumer often requires rational as well as emotional reasons for making a purchase.

7.4 Segmentation

Most of the above analyses can be considerably improved if we precede them by a segmentation analysis. The rationale for this is that an average ideal point or average determinant attribute is often misleading. There are usually market segments, each having its own ideal point and determinant attributes.

The best way to identify these segments is to perform a k-means type cluster analysis. This clustering should be based on general lifestyle statements as well as attitudinal statements that refer particularly to the product category in question. We have also found that using the importance ratings produces nice segments, each of which regards a bundle of attributes as important. These segments will then have very distinct ideal points. One should also plot the brand positions for each segment on the same map to establish whether the different segments have different perceptions of the brands.

While we will not show the maps and ideals for the segments in our clothing store example, we simply mention the five segments which we identified, namely: "Trendies," "Rationals," "Sophisticates," "Unfashionables," and "Cheap Trendies."

8. CONCLUSIONS

- Different data transformations and weightings, rather than particular statistical techniques, have a greater influence on the brand map produced.

- Techniques such as rotation and supplementary point fitting are not exclusive to any particular technique.

Finally, the method illustrated of producing importance weighted spaces shows a lot of promise, but needs to be tested on a lot more surveys in practice. An important consideration is also whether the results are understandable and can be effectively communicated to users of research.

7.3 Determinant attributes

The importance-weighted space and the ideal point directions have given us a good indication of the determinant attributes (see Section 6.6 and Figure 8).

The thread that runs through the all the analyses is that there are certain attributes which emerge from direct elicitation of importance and that these are not always the same as attributes whose importance is inferred from other data (e.g., preference). The first type are often the more rational type attributes (e.g., good value), while the latter are often more "image" or emotional type attributes.

In our view, this is not a major problem as long as we are able to identify the two types. In terms of the theories of consumer behavior a consumer often requires rational as well as emotional reasons for making a purchase.

7.4 Segmentation

Most of the above analyses can be considerably improved if we precede them by a segmentation analysis. The rationale for this is that an average ideal point or average determinant attribute is often misleading. There are usually market segments, each having its own ideal point and determinant attributes.

The best way to identify these segments is to perform a k-means type cluster analysis. This clustering should be based on general lifestyle statements as well as attitudinal statements that refer particularly to the product category in question. We have also found that using the importance ratings produces nice segments, each of which regards a bundle of attributes as important. These segments will then have very distinct ideal points. One should also plot the brand positions for each segment on the same map to establish whether the different segments have different perceptions of the brands.

While we will not show the maps and ideals for the segments in our clothing store example, we simply mention the five segments which we identified, namely: "Trendies," "Rationals," "Sophisticates," "Unfashionables," and "Cheap Trendies."

8. CONCLUSIONS

- Different data transformations and weightings, rather than particular statistical techniques, have a greater influence on the brand map produced.

- Techniques such as rotation and supplementary point fitting are not exclusive to any particular technique.

- Brand maps can be used as a simple graphical display of brand-attribute data or as a more analytical technique, by incorporating attribute importances and brand preferences.

- The use of respondent-level data is theoretically preferable to condensed data, but in practice the resulting displays are not very different.

- Using respondent-level data, we are able to plot individual respondent points for each brand, individual ideal points, and confidence regions for the brands.

- Brand-attribute associations are usually sufficient for producing brand maps.

- Collecting negative as well as positive attribute association data leads to less ambiguity in brand maps.

- Preceding brand mapping by segmentation enriches the interpretation of the maps.

- Importance-weighted spaces are easily handled by correspondence analysis - only a simple transformation of the data is required before applying the usual analysis.

- Correspondence analysis provides a single framework within which brand mapping and ideal point objectives may be accomplished.

REFERENCES

- Carroll, J. Douglas, Paul E. Green & Catherine M. Schaffer (1986), "Interpoint distance comparisons in correspondence analysis," Journal of Marketing Research 23 (August), 271-80.

- Carroll, J. Douglas, Paul E. Green & Catherine M. Schaffer (1987), "Comparing interpoint distances in correspondence analysis: a clarification," Journal of Marketing Research 24 (November), 445-50.

- Greenacre, Michael J. (1984), Theory and Applications of Correspondence Analysis. London: Academic Press.

- Greenacre, Michael J. (1986), "SimCA: a program to perform correspondence analysis," American Statistician 51, 230-231.

- Greenacre, Michael J. (1989), "The Carroll-Green-Schaffer scaling in correspondence analysis: a theoretical and empirical appraisal," Journal of Marketing Research 26 (August), to appear.

- Greenacre, Michael J. & Leslie G. Underhill (1982), "Scaling a data matrix in low-dimensional Euclidean space. In Topics in Applied Multivariate Analysis (ed. D. M. Hawkins), pp. 183-268. Cambridge, U.K.: Cambridge University Press.

- Greenacre, Michael J. & Trevor J. Hastie (1987), "The geometric interpretation of correspondence analysis," Journal of the American Statistical Association 82 (June), 437-47.

DISCRIMINANT VERSUS FACTOR BASED PERCEPTUAL MAPS: PRACTICAL CONSIDERATIONS

Thomas L. Pilon, IntelliQuest, Inc.

INTRODUCTION

There are numerous approaches to perceptual mapping (for an excellent summary see Neal). One useful way to classify these approaches is by the type of data used:

Similarities Data	-	result from direct judgment about similarities among products,
Attribute Data	-	result from respondents' rating of products on attributes; similarities are inferred from differences in their ratings, and
Contingency Data	-	are obtained from contingency table output.

When perceptual maps are based on similarities data, the similarities are decomposed to form a map. Historically, this "decomposition approach" was most widely used. Various mainframe software packages, mostly from Bell Labs, were used to generate the maps. Particularly popular software packages were KYST, INDSCAL, and PREFMAP.

More recently, however, researchers have preferred to compose maps by combining ratings of products on attributes since:

- similarities data are relatively inefficient to collect,
- maps based on the decomposition approach have been found to be relatively difficult to interpret, and
- the computational procedures are relatively less robust than the computational procedures used in the "composition approach."

The two computational methods most widely used in the product attribute composition approach are Factor Analysis and Multiple Discriminant Analysis.

A third approach, Correspondence Analysis (also know as dual scaling), utilizes contingency table data. The primary benefit of Correspondence Analysis is that it can be used on nominal or ordinal data, whereas the approaches described above require intervally scaled data. This allows the researcher to generate maps from data that are easier to collect (or have already been collected in a form not appropriate for the other approaches). Another advantage of Correspondence Analysis is that the maps are very easy to interpret since they have no vectors; each item and each scale is represented as a single point. This advantage is somewhat offset by the fact that the maps contain less information.

OVERVIEW

This paper compares the two computational methods that are most widely used for the analysis of attribute data (composition approach): Factor Analysis and Multiple Discriminant Analysis. The two methods will be compared on the following dimensions:

- nature of methods
- types of variance used in calculations

- nature and interpretability of factors or dimensions extracted
- number of factors or dimensions extracted
- sensitivity of extracted factors or dimensions
- available tests of significance
- popularity

Exhibit A contains a summary of the comparisons.

NATURE OF METHODS

The goal of Factor Analysis is to generate dimensions that maximize interpretability and explained variance. The first factor equation is that linear combination that explains the maximum amount of correlation among the attributes. Factor Analysis is an interdependence procedure; the various differences in product ratings are ignored until after the factor equations are derived.

Factor Analysis is a multivariate statistical technique that is concerned with the identification of structure within a set of observed attributes. Its appropriate use involves the study of interrelationships among attributes in an effort to find a new set of factors, fewer in number than the original attributes, which express that which is common among the original attributes. Factor Analysis establishes dimensions within the data and serves as a data reduction technique. "Three general functions may be served by Factor Analysis:

1. The number of variables for further research can be minimized while the amount of information in the analysis is maximized. The original set of variables can be reduced to a small set which accounts for most of the variance of the initial set.

2. When the amount of data is so large as to be beyond comprehension, Factor Analysis can be used to search data for qualitative and quantitative distinctions.

3. If a domain is hypothesized to have certain qualitative and quantitative distinctions, Factor Analysis can test this hypothesis... Much of the expressed dissatisfaction with Factor Analysis can be attributed to its use for purposes other than those stated above." (Stewart)

The goal of Discriminant Analysis is to generate dimensions that will discriminate or separate the products as much as possible. The first discriminant equation is that linear combination that explains the maximum amount of differences between the products. Discriminant Analysis develops axes for the perceptual map in sequential order, starting with the feature or group of similar features that best distinguish among the products being evaluated. The relative importance (size) of each discriminant function is given in quantitative terms. Discriminant Analysis highlights those attributes that discriminate between products.

"The basic premise for positioning by Discriminant Analysis is similar to that for non-metric scaling of object similarities data: the greater the perceived difference among objects on a particular feature, the more important that feature is in estab- lishing the perceptual space for these objects." (Meyers & Tauber)

Exhibit A

ATTRIBUTE-RATING BASED MAPPING

Factor Analysis (FA)

Characteristics

- is an interdependence procedure;
  - thus the various differences in product ratings are ignored until after the factor equations are derived
  - the first factor equation is that linear combination that explains the maximum amount of correlation between the attributes
- uses only total variation in calculations;
- FA extracts factors tending to emphasize those attributes that prompt agreement about the meaning of the attributes;
- produces dimensions that are purer and easier to interpret;
- uses more attributes in the dimensions;
- extraction of factors is highly sensitive to the number of correlated attributes;
  - the addition or deletion of an attribute may dramatically alter the dimensionality of the derived space
  - a single attribute that may be considered extremely important and dominating the selection of products, like safety, may not show up as a dimension on a map, simply because it is not correlated to any other measures
- no tests of significance are available;
- seldom used as an attribute-rating based mapping technique, although in the '70's it was the preferred mapping procedure.

Discriminant Analysis (DA)

Characteristics

- highlights (is limited to) those attributes that discriminate between products;
  - the first discriminant equation is that linear combination that explains the maximum amount of difference between the products,
- uses both among-brand and within-brand variation in calculations;
- extracts dimensions based on those attributes about which respondents agree in their perceptual positioning of the brands;
- is more powerful at capturing information in fewer dimensions;
- uses fewer attributes in the dimensions;
- extractions of dimensions is highly sensitive to the products included in the space;
  - the deletion (or addition) of a product may effect the dimension ality of the derived space, although the relationships of the remaining (previous) products would not change
  - will not display on the map attributes that may be extremely important, even dominating product choice, but that do not differentiate between products
- tests of significance are available;
- most popular attribute-rating mapping technique in the '80's.

TYPES OF VARIANCE USED IN CALCULATIONS

"It is also possible to construct perceptual spaces using Factor Analysis. In this case, the dimensions would be defined in terms of linear combinations that maximize the sum of the within brand variance (i.e. disagreement about the brand) plus the among brand variance (i.e. differences across brands). While maximizing explained variance is usually a good property, it does not seem logical that one would want a linear combination that maximized the within brand variance since it is a measure of confusion." "However from a practical standpoint, the between brand variance is usually much bigger than the within variance, so the two procedures give similar answers in many cases. Still Discriminant Analysis seems to offer the conceptually superior approach. " (Moore & Pessemier)

NATURE & INTERPRETABILITY OF FACTORS OR DIMENSIONS EXTRACTED

Factor Analysis groups together attributes or features that are seen as similar (i.e. rated about the same) by respondents. The resulting groupings are descriptive adjectives that mean about the same thing to people; hence, Factor Analysis simply performs an exercise in semantics by identifying groups of similar statements. The size of each grouping depends only on the number of feature statements that are seen as similar and how similar they are seen to be. In turn, the number of similar statements that are included for a particular grouping depends only on the very arbitrary choice of the researcher. Therefore, it is actually true that the axes of a Factor Analysis positioning map can be determined while designing the questionnaire by deciding how many statements about a given aspect of a product are to be rated by respondents.

The extraction of factors is sensitive to the number of correlated attributes. In addition, extraction of factors is dependent on the intercorrelations among attributes, and does not optimize the separation between products, like Discriminant Analysis. Furthermore, a single attribute that may be considered extremely important and dominating the selection of products may not show up as a dimension on a map, simply because it is not correlated with any of the other attributes.

"Factor Analysis produces vectors or axes by looking for groups of attributes that are similar in meaning to the respondent: for example, nutritious, good for teeth, good for complexion in the case of snack foods. This means that even features that are most important in determining choice behavior would certainly be overlooked if they were relatively independent of other features being rated; that is they would not statistically group with other features and therefore would not emerge as a factor." "The size or importance of a dimension extracted by Factor Analysis is based only on how many other features are judged to be similar and how similar these other features are - the more that are similar, the greater the apparent importance of the vector or axis. "It might caustically be observed that if degree of importance of factors extracted in conventional Factor Analysis turned out to be the same as their actual importance in influencing choice decisions, it would be only by accident!" (Meyers & Tauber)

In contrast to Factor Analysis, Discriminant Analysis does not require groups of similar features to establish an axis. A single isolated feature will be selected if it discriminates better among products than any of the groupings. The extent to which an attribute will be an important contributor toward a dimension depends upon the extent to which there is a perceived difference among the products on that attribute. Discriminant Analysis will not display on the map attributes that may be extremely important, even those attributes that dominate product choice, if they do not differentiate among products.

NUMBER OF FACTORS OR DIMENSIONS EXTRACTED

Discriminant Analysis is more powerful at capturing information in fewer dimensions since it is limited to dimensions that distinguish among products.

Although Hauser and Koppleman state that "Factor Analysis should use more attributes than Discriminant Analysis in the dimensions and therefore produce richer solutions," it does not necessarily follow that the use of more attributes produces richer solutions if the attributes are providing the same (highly correlated) information.

SENSITIVITY OF EXTRACTED FACTORS OR DIMENSIONS

In Factor Analysis, the extraction of factors is highly sensitive to the number of correlated attributes. The addition or deletion of an attribute may dramatically alter the dimensionality of the derived space.

In Discriminant Analysis the extraction of dimensions is highly sensitive to the products included in the space. The deletion (or addition) of a product may affect the dimensionality of the derived space.

AVAILABLE TESTS OF SIGNIFICANCE

An advantage of Discriminant Analysis over Factor Analysis is that a test of significance is available. Discriminant Analysis provides statistical tests of significance of the differences among all products to show the extent to which all products differ in perceptual space, and a test for the difference between any two products.

POPULARITY

"Factor Analysis is seldom used as a mapping procedure in today's applied marketing research field, although in the 1970's it was the preferred mapping procedure among many applied researchers." "Discriminant Analysis is still the most popular algorithm in use for today for applied multivariate mapping." (Neal)

EMPIRICAL OBSERVATIONS FROM THE LITERATURE

"A comparison of different methods of building joint spaces indicates significant differences both across product classes and modeling techniques in the ability to predict preferences for validation brands. Perceptual spaces constructed with Discriminant Analysis combined with vector representations of preference gave the best predictions across the product classes analyzed." (Moore)

"...only two studies have compared the predictive power of these two compositional methods. Both found that Factor Analytic spaces were superior to those constructed using Discriminant Analysis. This is somewhat surprising as theoretical reasoning would suggest that Discriminant Analysis should give a better representation of perception that should, in turn, lead to better predictions. With a goal of eliminating this discrepancy between theoretical arguments and empirical findings, the present study extends the previous studies in a number of important ways: 1) in addition to the vector model of preference previously studied, an ideal point model is also compared, 2) a greater degree of preferential heterogeneity and different type of perceptual heterogeneity is allowed, 3) models are compared as to their ability to predict preferences for a holdout brand rather than preference for brands used to calibrate the model, and 4) the comparisons are carried out across four product or service categories. When these extensions are considered, it is found that discriminant spaces do, in fact, give superior predictions." (Moore)

Huber and Holbrook found that the use of Discriminant Analysis, as opposed to Factor Analysis, produced dimensions that were more objective in reflecting homogeneous perceptions across consumers. In addition, they found that Discriminant Analysis solutions were more likely to be actionable - that is, they were more likely to indicate specific actions the manufacturer must take to build such a product. Conversely, they found that Factor Analysis often provided solutions that were more sensitive to the meaning of attributes. They suggested that such analyses might be more appropriate in, say, an advertising study where managerial attention is focused on semantic labeling instead of objective brand differences. "Principal Components Analysis tends to orient a space to dimensions that have high variance both within and between the objects. Hence the semantic meaning of the adjectives with high loadings is very stable. This kind of analysis is particularly relevant to formulating a communications strategy in which the linguistic relations between attributes are critical." (Huber and Holbrook) They went on to say that Discriminant Analysis is appropriate where one is concerned with product design attributes that can be clearly and unequivocally perceived by consumers.

ADDITIONAL EMPIRICAL OBSERVATIONS

To supplement the above discussion, several perceptual maps were created. The database is proprietary and cannot be described, but an understanding of the study from which the data were borrowed is not essential to demonstrate and compare the techniques.

Exhibits B, E, H and I were generated by Sawtooth Software's Adaptive Perceptive Mapping System (APM), which utilizes the Discriminant Analysis computational method (Johnson, 1987). Exhibits C, F and G were created by Factor Analyzing the product ratings on the attributes. The space was then varimax rotated. The product locations in the space were derived by plugging the product ratings on each attribute into the factor equations and calculating the X and Y coordinates. Exhibit D was generated by Factor Analyzing the attribute importance ratings. The space was also varimax rotated. The product locations in the space were derived in the same manner. In all cases, only the first two discriminant functions or factors were significant.

Exhibits B, C, D were developed from the complete data set. The discriminant space results in a relatively broader dispersion of the products. The two-factor spaces are quite similar to one another (if one visualizes a 120 degree spatial rotation of Exhibit C). A closer examination reveals that the factor spaces and the discriminant space would be somewhat similar if one were to drop a perpendicular from each product point to an axis that includes both product G and the origin.

Exhibit E was generated by dropping one attribute ("integrated") from Exhibit B. The "integrated" attribute was selected because it was the attribute most highly correlated with the other attributes. As one can see, Exhibits B and E are quite similar. Various other attributes (including "large," "open," and "proprietary") and combinations of attributes were dropped with minimal changes to the resulting maps.

Exhibit F was created by dropping "integrated" from Exhibit C. The resulting maps are again quite similar (again, if one visualizes a 120 degree spatial rotation of Exhibit C). Similar results (see Exhibit G) were obtained when "large," "open," and "proprietary" were dropped from the factor space in Exhibit C.

Exhibits H and I demonstrate the effect of dropping products from a Discriminant space. Although dropping product G (Exhibit I) results in a radical change in the space, the relative position of the remaining products is fairly well preserved.

While these Exhibits demonstrate some interesting results, one must be cautioned against making general or conclusive statements based on the analysis of one database.

SUMMARY

In reference to when researchers first became discontented with similarities data-based maps, Meyers and Tauber attempted to explain the early use of Factor Analysis as an alternative mapping technique. "One of the most obvious alternatives was Factor Analysis, a technique that was both widely understood and easily applied by most investigators. It was readily adaptable for MDS purposes, even though it was not formally a part of that body of technology. Moreover, the output format was for all practical purposes indistinguishable. And, it required input data that were often more easily obtained and better understood by most market planners. Unfortunately, it also contained conceptual weaknesses that were not readily apparent even to many researchers, much less operating marketing executives." "Nonetheless, Factor Analysis can be defended on grounds that it is a legitimate data-reduction technique for finding the basic dimensions underlying a group of attributes that characterize objects under investigation. Whether the results are meaningful from a market planning standpoint is another matter." (Meyers & Tauber)

Discriminant Analysis is probably the best technique for developing a perceptual map when metric ratings of product characteristics on interval scales are desired. There is an assumption, of course, about a perceptual space based on perceived similarities and differences among products being evaluated: The greater the perceived difference among products on a given feature, the more important that feature is in establishing a perceptual space. Some features of a product may not be perceived to differ among existing brands to any meaningful extent; for example, noise level of personal computers. Noise level would not emerge as an axis of a discriminant map; thus, it could easily be overlooked in designing new computers. Yet any personal computer that was not perceived as being relatively quiet by consumers would probably fail.

(SUMMARY continued on page 181)

EXHIBIT B

DISCRIMINANT ANALYSIS SPACE

ALL ATTRIBUTES - ALL PRODUCTS

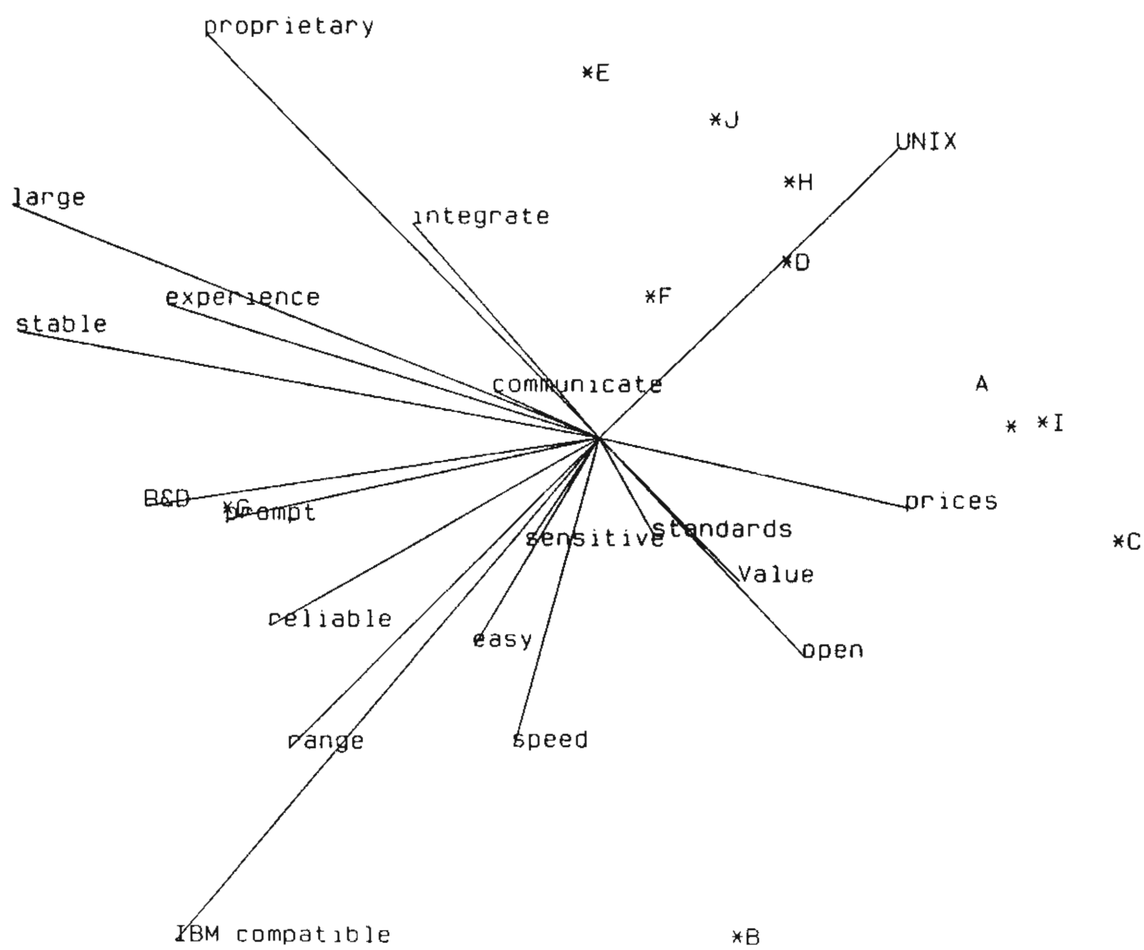


EXHIBIT C

FACTOR ANALYSIS SPACE BASED ON PRODUCT RATINGS

ALL ATTRIBUTES - ALL PRODUCTS

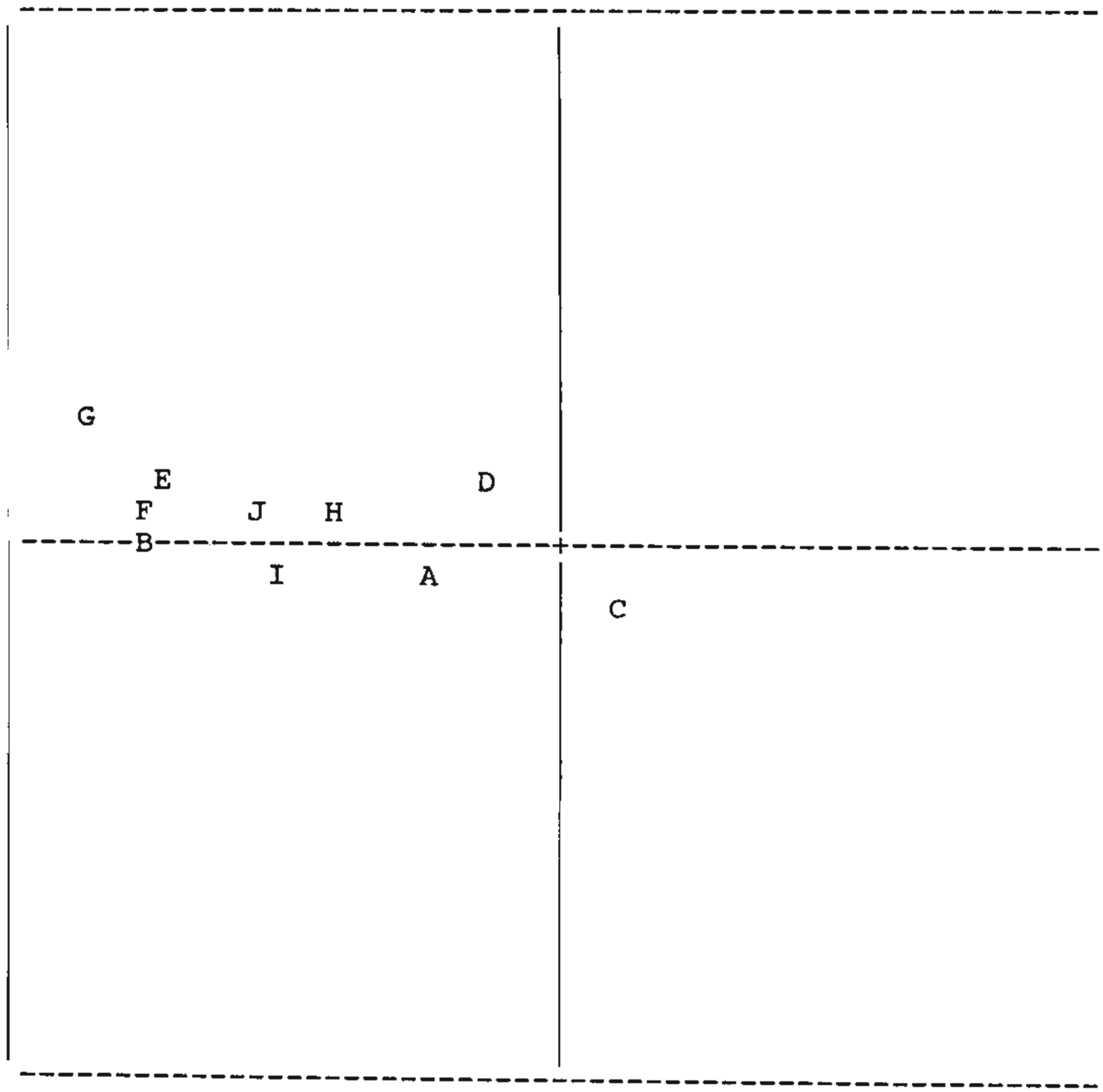


EXHIBIT D

FACTOR ANALYSIS SPACE BASED ON IMPORTANCE RATINGS

ALL ATTRIBUTES - ALL PRODUCTS

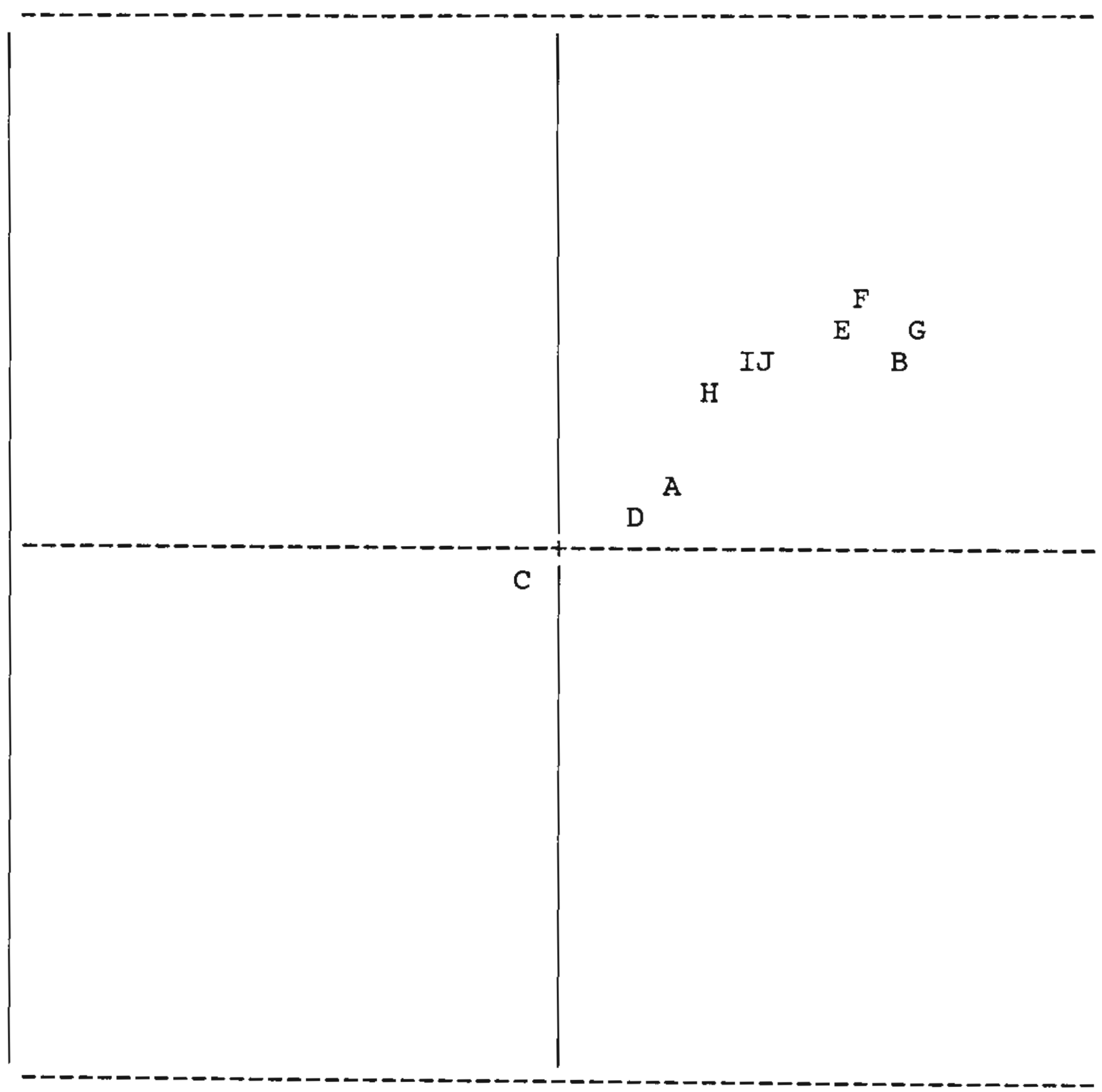


EXHIBIT E

DISCRIMINANT ANALYSIS SPACE

ALL BUT ONE ATTRIBUTE - ALL PRODUCTS

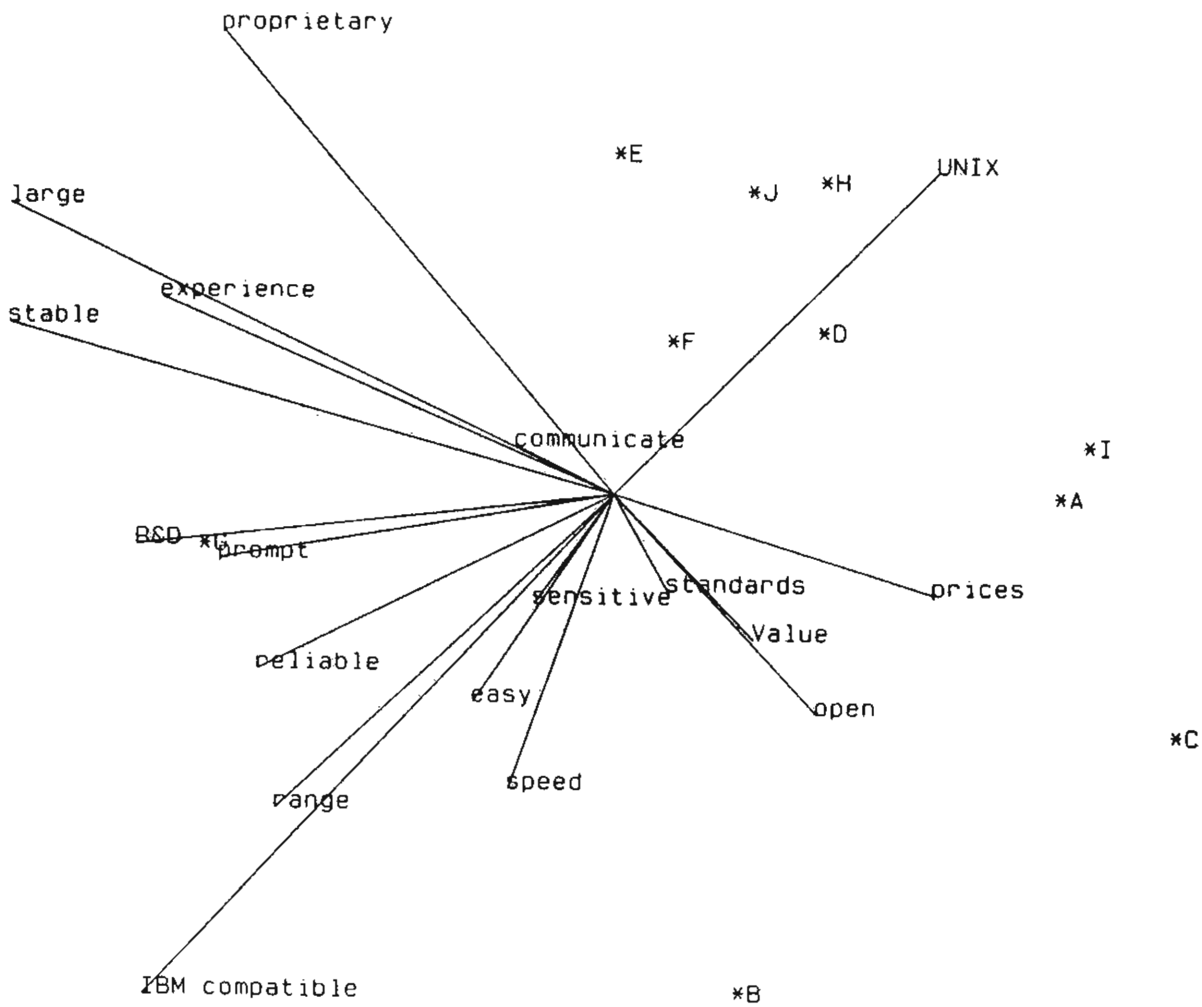


EXHIBIT F

FACTOR ANALYSIS SPACE BASED ON PRODUCT RATINGS

ALL BUT ONE ATTRIBUTE - ALL PRODUCTS

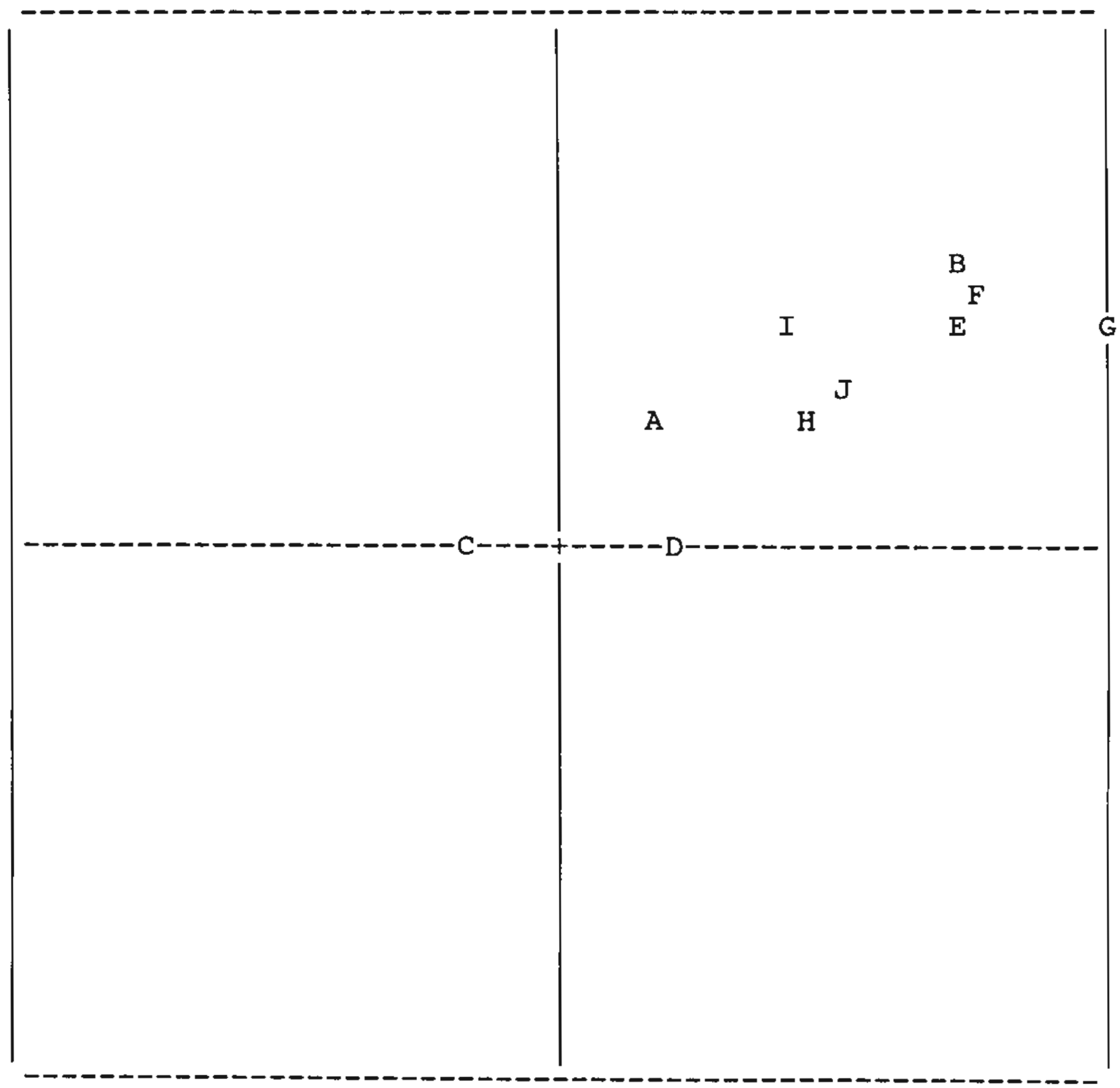


EXHIBIT G

FACTOR ANALYSIS SPACE BASED ON PRODUCT RATINGS

ALL BUT FOUR ATTRIBUTES - ALL PRODUCTS

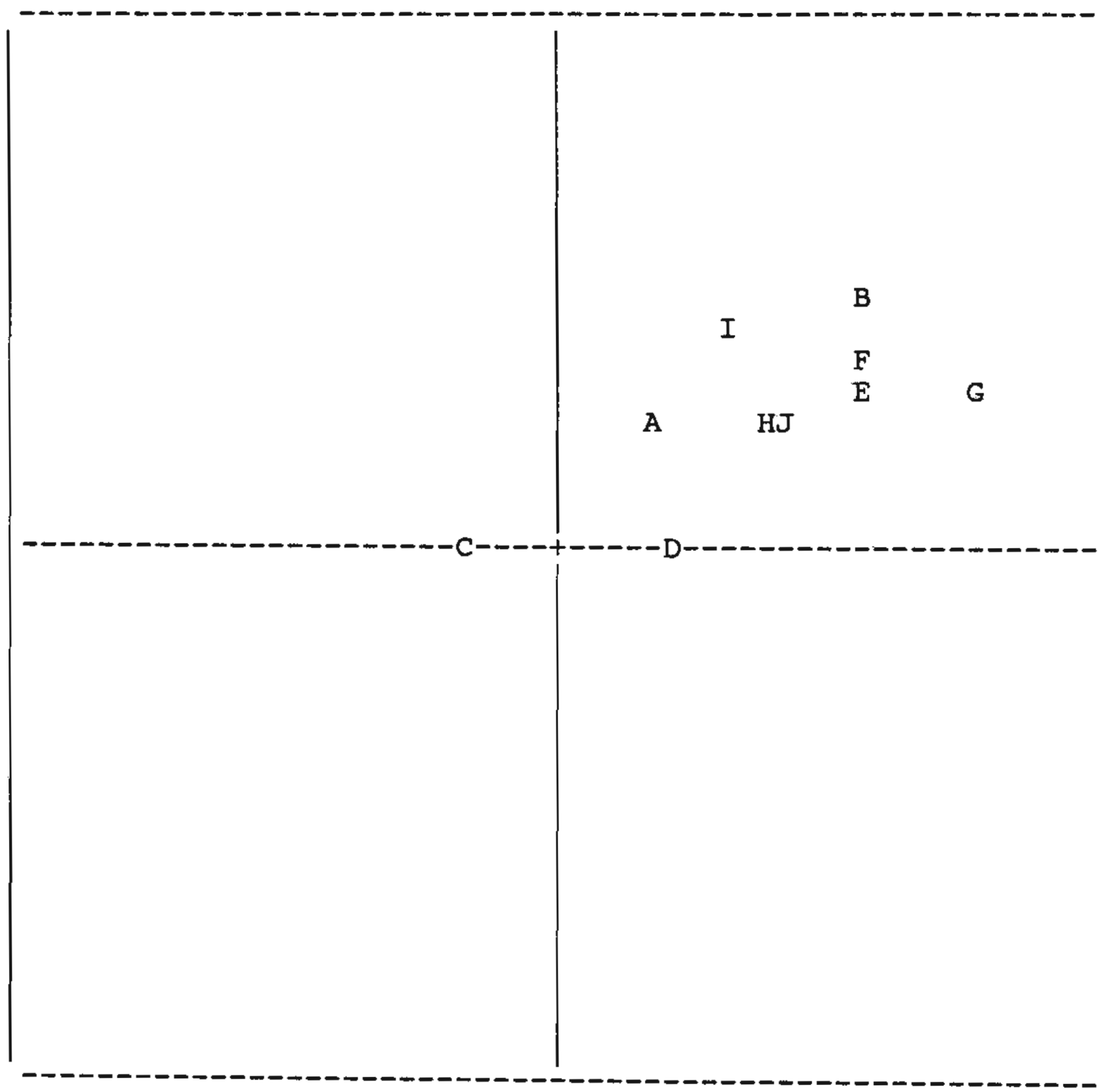


EXHIBIT H

DISCRIMINANT ANALYSIS SPACE

ALL ATTRIBUTES - ALL BUT PRODUCT J

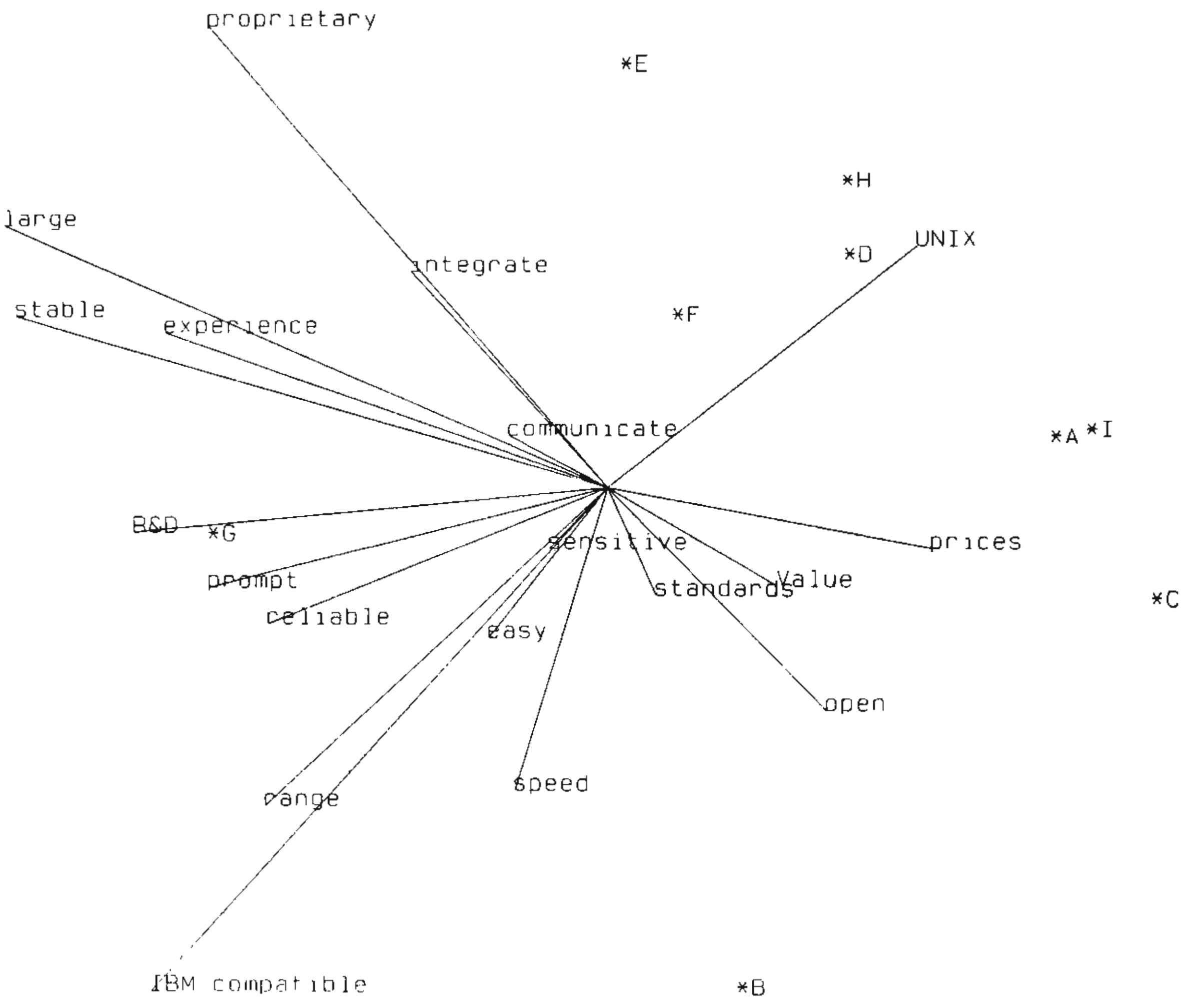
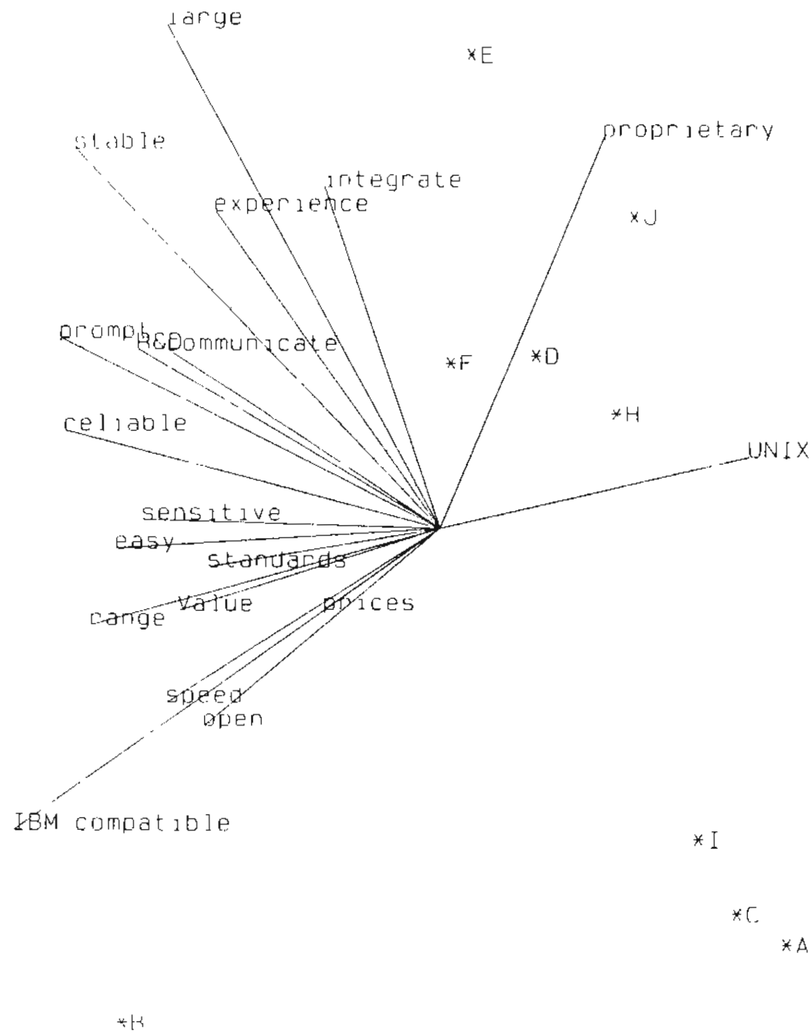


EXHIBIT I

DISCRIMINANT ANALYSIS SPACE

ALL ATTRIBUTES - ALL BUT PRODUCT G



This latter point is a problem common to all types of perceptual maps that focus on the major differences among competing products or services. These maps appear to represent to the marketer (even though they may not actually do so) that features that do not distinguish much among existing offerings are of little importance. Also, these maps appear to represent that the more a particular feature distinguishes among existing offerings, the more important it is for new product and promotion planning purposes. Neither of these appearances may be true.

Finally, since perceptual maps are devices for visual comprehension, most researchers prefer the Discriminant Analysis approach since it is vastly preferable to have only two, or at most three, dimensions. Since perceptual maps are often used to communicate strategic concepts to non-researchers at high management levels, the complex interpretation of higher-dimensional spaces that often result from the Factor Analysis approach is usually not feasible.

REFERENCES

Aaker, David, "Multidimensional Scaling," in Multivariate Analysis in Marketing, 2nd ed., Scientific Press, Palo Alto, CA, 185- 191.

Hauser, John R. and Frank S. Koppleman, "Alternative Perceptual Mapping Techniques: Relative Accuracy and Usefulness," Journal of Marketing Research, Volume XVI (November 1979), 495-506.

Huber, Joel and Morris B. Holbrook, "Using Attribute Ratings for Product Positioning: Some Distinctions Among Compositional Approaches," Journal of Marketing Research, Volume XVI (November 1979), 507-16.

Johnson, Richard M., "Multiple Discriminant Analysis: Marketing Research Applications," in Multivariate Methods for Market and Survey Research, Jadgish Sheth, ed., 1971, 65-82.

Johnson, Richard M., "Adaptive Perceptual Mapping," Proceedings of the Sawtooth Software Conference on Perceptual Mapping, Conjoint Analysis, and Computer Interviewing, 1987, 143-158.

Moore, William L., "Predictive Power of Joint Space Models Constructed with Composition Techniques," Journal of Business Research, Volume 10, 1982, 217-236.

Moore, William L. and Edgar Pessemier, Computer Aided MARKETING PLANNING, DRAFT: 1 September 1988.

Myers, James H. and Edward Tauber, Market Structure Analysis, American Marketing Association, 1977.

Neal, William D., "Overview of Perceptual Mapping," Proceedings of the Sawtooth Software Conference on Perceptual Mapping, Conjoint Analysis, and Computer Interviewing, 1988, 151-163.

Simmie, Patricia, "Alternative Perceptual Models: Reproducibility, Validity, and Data Integrity," Proceedings of the American Marketing Association Educators Conference, August, 1978.

Stewart, David, "The Application and Misapplication of Factor Analysis in Marketing Research," Journal of Marketing Research, Volume XVIII (February 1981), 51-62.

The Manager versus. the Customer: A Comparison of Values

Richard B. Ross and *Larry G. Gulledge*
Elrick and Lavidge, Inc.

For years, fast food chains have measured individual unit performance in areas of quality, service, and cleanliness. Many operational and training departments of successful chains routinely monitor individual transactions at their restaurants in an effort to align performance with customer expectations and company standards. It's not uncommon for management compensation to be directly tied to the scores received during the monitored transactions. Generally speaking, this industry knows its customers. And, successful chains are constantly striving to motivate their employees to meet customer expectations and to exceed them, if possible.

Considerable marketing research has been devoted to determining what features of a fast food restaurant are important to customers and then tracking performance relative to these features using various data collection methodologies. Administering rating scales over the telephone is one commonly used methodology, although exit interviewing, comment cards, and mystery shopping techniques are also used.

For the purposes of this paper, however, we have attempted to measure the differences that exist between managers of a fast food restaurant and customers: Do these managers evaluate the importance of attributes in the same way as their customers evaluate them? To this end a conjoint study was designed that exposed the respondents (both managers and customers) to different experiences that required them to select one situation over another. The following thirteen attributes were included, with all but one having three levels:

ATTRIBUTE #1

10 Restaurant had a very clean exterior.

20 Restaurant had a somewhat clean exterior.

30 Restaurant had a not very clean exterior.

ATTRIBUTE #2

10 Restaurant was very clean inside.

20 Restaurant was somewhat clean inside.

30 Restaurant was not very clean inside.

ATTRIBUTE #3

10 Restrooms were very clean.

20 Restrooms were somewhat clean.

30 Restrooms were not very clean.

ATTRIBUTE #4

10 Food was very attractive looking.

20 Food was somewhat attractive looking.

30 Food was not very attractive looking.

ATTRIBUTE #5

10 Food was very tasty.

20 Food was somewhat tasty.

30 Food was not very tasty.

ATTRIBUTE #6

10 Temperature of food was just right.

20 Temperature of food was a little off.

30 Temperature of food was too hot or too cold.

ATTRIBUTE #7

10 Meal was an exceptional value.

20 Meal was an average value.

30 Meal was a below average value.

ATTRIBUTE #8

10 Location of restaurant was very convenient.

20 Location of restaurant was somewhat convenient.

30 Location of restaurant was not very convenient.

ATTRIBUTE #9

10 Restaurant workers were very friendly.

20 Restaurant workers were somewhat friendly.

30 Restaurant workers were not very friendly

ATTRIBUTE #10

10 Employees were above average in appearance.

20 Employees were average in appearance.

30 Employees were below average in appearance.

ATTRIBUTE #11

10 Service speed was fast.

20 Service speed was average.

30 Service speed was slow.

ATTRIBUTE #12

10 Order was served exactly right.

11 Order was served not exactly right.

ATTRIBUTE #13

10 Restaurant caters to young children.

20 Restaurant caters to adults.

30 Restaurant caters to both adults and young children.

This test used the ACA System (Adaptive Conjoint Analysis) published by Sawtooth Software. The ACA System has three parts:

- o An Interviewing Module for collecting respondent data
- o A Utility Calculator for estimating each respondent's utilities for product features
- o A Market Simulator for testing alternative product scenarios.

The interview was structured by the ACA System. We simply entered the list of attributes and wrote the instructions that lead the respondent through the interview. The actual interview began by asking the respondent the following preference-ranking question:

> Type the number of your FIRST CHOICE,
>
> assuming everything else to be equal.
>
> 1. Restaurant caters to young children.
> 2. Restaurant caters to adults.
> 3. Restaurant caters to both adults and young children.

Once the first choice is entered the program asks the respondents to enter their next choice.

This attribute was included because it was discussed in great detail by focus group interview respondents during our exploratory study.

The next series of questions rates the importance of each of the thirteen attributes. An example is displayed below.

If two restaurant attributes were both acceptable in all other ways, how important would this difference be?

A: Restaurant had a very clean exterior.

versus

B: Restaurant had a not very clean exterior.

4 = Extremely important
(I could almost never accept B.)

3 = Very important (B would have to be
outstanding in other ways.)

2 = Somewhat important (But I would not
base my decision on this.)

1 = Not important at all.

To answer type a number from the scale.

The Importance Ratings together with the Preference Rankings let us estimate an initial set of utilities for the respondent before going into the Paired Concept section.

Each respondent now sees a series of paired concepts that are unique to him. They are constructed based on an estimate of his utilities updated after each paired question using a Bayesian algorithm.

The ACA System allows you to construct pairs having two to five attributes per concept. We used three per concept as displayed below.

Strong Prefer Top	Meal was an exceptional value.
1	
2	Restaurant had a very clean exterior.
3	
4	Restaurant was not very clean inside.
--5--	OR
6	Meal was an average value.
7	
8	Restaurant had a somewhat clean exterior.
Strong Prefer Bottom	Restaurant was somewhat clean inside.

The final section of the interview is used to calibrate the respondent's utilities. The ACA System lets you ask up to 9 questions each, made up of as many as 9 attributes. We used 6 questions with 6 attributes as displayed below.

Considering restaurants at which you have eaten previously,

How likely would you be to eat there if you were eating out today?

9 = 90+%	
8 = 80%	Restaurant had a very clean exterior.
7 = 70%	Restaurant was somewhat clean inside.
6 = 60%	Restrooms were somewhat clean.
5 = 50%	Food was somewhat attractive looking.
4 = 40%	Food was somewhat tasty.
3 = 30%	Temperature of food was a little off.
2 = 20%	
1 = 10-%	

Type a number from 1 to 9 on the scale, or X to go back.

The interviews with fast food restaurant management were conducted at the individual restaurant locations using a Toshiba T300 lap-top computer. The administration time averaged 35 minutes with some respondents requiring as little as 25 minutes while others needed 45 minutes to complete the computerized personal interview. Administration time with customers was somewhat less, taking 15-20 minutes. The managers seemed to struggled more than customers with some of the trade-offs required during the paired concepts phase of the questionnaire.

Ordinary least squares regression is used to calculate utilities for each respondent and to scale the utilities.

Table I below displays the utilities coefficients and their range for each attribute by respondent type. Table II displays the rank order of attributes based on the range of utility coefficients for both managers and customers.

TABLE I: UTILITY COEFFICIENTS

ATTRIBUTE		MANAGERS		CUSTOMERS
Very Clean Inside		85		74
Somewhat Clean Inside		45		45
Not Clean Inside		4		2
RANGE	81		72	
Very Attractive Food		104		119
Somewhat Attractive Food		63		72
Not Attractive Food		0		0
RANGE	104		119	
Very Tasty Food		97		98
Somewhat Tasty Food		55		57
Not Very Tasty Food		3		0
RANGE	94		98	
Temperature Just Right		90		100
Temperature a Little Off		47		54
Too Hot or Too Cold		4		0
RANGE	86		100	
Exceptional Value		124		128
Average Value		84		67
Below Average Value		3		0
RANGE	121		128	
Very Friendly Workers		95		100
Somewhat Friendly Workers		37		37
Not Using Friendly Workers		8		1
RANGE	87		99	
Employees Were Above Average Appearance		101		86
Employees Were Below Average Appearance		63		55
Employees Were Below Average Appearance		0		0
RANGE	101		86	
Fast Service		82		79
Average Speed		50		48
Slow Service		1		4
RANGE	81		75	

ATTRIBUTE	MANAGERS	CUSTOMERS
Order Served Exactly Right	108	91
Order Served Not Exactly Right	75	57
RANGE	33	34
Caters to Adults/Young Children	0	0
Caters to Young Children	101	92
Caters to Adults	57	60
RANGE	101	92
Very Clean Exterior	2	1
Somewhat Clean Exterior	105	106
Not Very Clean Exterior	61	68
RANGE	103	105
Very Clean Restrooms	5	1
Somewhat Clean Restrooms	98	118
Not Very Clean Restrooms	0	0
RANGE	98	118
Very Convenient Location	22	11
Somewhat Convenient Location	26	44
Not Very Convenient Location	77	77
RANGE	55	66

TABLE II: RANGE OF UTILITY COEFFICIENTS RANKED BY ATTRIBUTE IMPORTANCE

CUSTOMERS		MANAGERS	
Value of Meal	128	Value of Meal	121
Attractive Food	119	Attractive Food	104
Clean Restrooms	118	Clean Exterior	103
Clean Exterior	105	Employee Appearance	101
Temperature of Food	100	Caters to Children	101
Friendly Workers	99	Clean Restrooms	98
Tasty Food	98	Tasty Food	94
Caters to Children	92	Friendly Workers	87
Employee Appearance	86	Temperature of Food	86
Service Speed	75	Service Speed	81
Clean Inside	72	Clean Inside	81
Convenient Location	66	Convenient Location	55
Order Seemed Right	34	Order Seemed Right	33

The sample size for this study is too small to be statistically valid but it does enable us to make some empirical observations that can be addressed more quantitatively in future research.

The utility coefficients displayed in Table II suggest that both customers and managers place the most importance on the value of the meal and the attractiveness of the food while placing the least importance on the accuracy of the order and the convenience of the location. The most obvious difference between managers and customers appears to be the importance of the restaurant having clean restrooms, although both groups recognize unclean restrooms as being unacceptable. Another difference is the appearance of the restaurant employees. Managers placed more importance on this attribute than did customers.

The low importance rating of service speed to both managers and customers is interesting to me. Service speed is a feature regularly monitored by corporate management and considered very important. Possibly, since our topic for this study was fast food hamburger restaurants, service speed is taken for granted or is assumed to be relatively fast even at the slowest fast food restaurant. Also interesting is the lower positioning of interior cleanliness with both managers and customers, while both groups place higher importance on the exterior cleanliness.

Overall, some differences do exist. But, 7 of the 13 attributes are ranked in the same order of importance by both customers and managers.

Unit managers, assistant managers, and supervisors of one fast food hamburger chain were interviewed for this test. As mentioned earlier, the manager interviews were conducted at the individual unit restaurants in training nooks or offices located in areas designated for employees only. Unfortunately, most of the units did not have an area available for customer interviews. Therefore, the customers were interviewed off-site. Each cooperating customer respondent was given a coupon worth a free meal at any of the client's restaurants.

Once the conjoint interview was completed, each respondent was then interviewed regarding his perception of how well the restaurant chain was performing relative to each attribute. Table III below displays the mean scores for each attribute on a 3-point scale. Also shown is an index that was calculated to provide easier interpretation.

TABLE III: PERCEPTION OF PERFORMANCE

	--MANAGERS --		--CUSTOMERS --	
	MEAN	INDEX	MEAN	INDEX
TASTY FOOD	2.85	95	2.44	81
ATTRACTIVE FOOD	2.80	93	2.56	85
ACCURACY OF ORDER	2.70	90	2.81	94
LOCATION	2.60	87	2.56	85
RESTROOM CLEANLINESS	2.45	82	2.38	79
EXTERIOR CLEANLINESS	2.40	80	2.81	94
FOOD TEMPERATURE	2.37	79	2.75	92
INTERIOR CLEANLINESS	2.35	78	2.69	90
FRIENDLY EMPLOYEES	2.35	78	2.19	73
SERVICE SPEED	2.30	77	2.31	77
MEAL VALUE	2.25	75	2.19	73
EMPLOYEE APPEARANCE	2.20	73	2.13	71
TOTAL	2.47	82	2.49	83

To more easily compare the respondents' perceptions to the importance ratings, we calculated an index for each attribute by dividing the utility range by the highest range, to express importances as percentages, with the most important attribute at 100%. Table IV displays the comparison.

TABLE IV: INDEX OF IMPORTANCE VERSUS PERFORMANCE

	--- IMPORTANCE ---		--- PERFORMANCE ---	
ATTRIBUTE	CUSTOMERS	MANAGERS	CUSTOMERS	MANAGERS
MEAL VALUE	100	100	73	75
ATTRACTIVE FOOD	93	86	85	93
CLEAN RESTROOMS	92	81	79	82
CLEAN EXTERIOR	82	85	94	80
TEMPERATURE	78	71	92	79
FRIENDLY WORKERS	77	72	73	78
TASTY FOOD	77	78	81	95
CATERS CHILDREN	72	83	*	*
EMPLOYEE APPEARANCE	67	83	71	73
SERVICE SPEED	59	67	77	77
CLEAN INSIDE	56	67	90	78
LOCATION	52	45	85	87
ACCURACY OF ORDER	27	27	94	90

* 96% of customers and 95% of managers feel this chain caters to children or adults and children.

Some gaps between importance and perceptions of performance do exist. The value of the meal is one example. It is recognized as most important by both managers and customers, while both groups rate performance well below other attributes.

Cleanliness of the restrooms is an example of the customer placing high importance on this feature but rating the performance of this chain below the managers' perceptions of performance. It is interesting to note that during the interviews with customers, several commented on this feature saying that the fast food restaurant has replaced the service station as a place to stop while traveling - simply to use the restroom.

Overall, these customers report that this chain is serving a tasty, quality product, served accurately and in a clean environment. Although there is room for improvement related to restroom cleanliness, meal value, and friendliness of the employees, this chain appears to be doing a good job.

Managers also feel they are doing a good job. However, there are some interesting gaps between the managers and customers, such as the temperature of the food. Managers rated their performance on this feature considerably below their customers' perceptions. This is also true on exterior and interior cleanliness. Conversely, the managers rated their performance much higher than customers did on providing attractive tasty meals.

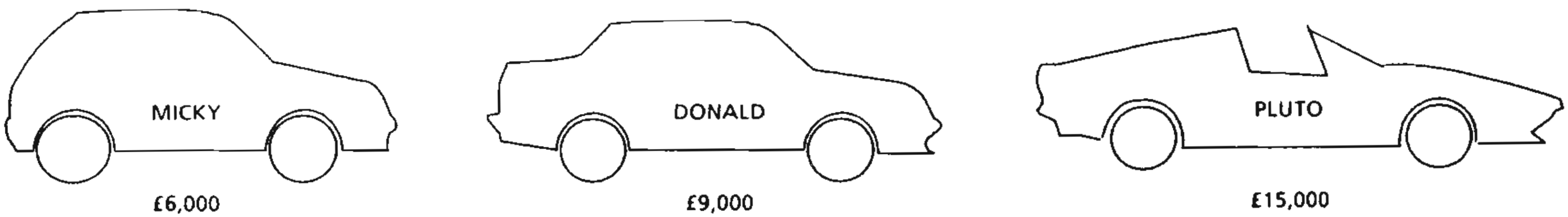
This test does suggest that the ACA system provides an excellent tool to evaluate the importance of attributes. When used in combination with performance measurement it is a sound method for measuring customer satisfaction. It would be especially useful if competitive performance measurements were also included so comparisons could be made and industry benchmarks could be established.

CONJOINT ANALYSIS ACROSS THE BUSINESS SYSTEM

Neil Allison
McKinsey & Company, Inc.

Traditionally we think of conjoint analysis as a tool for understanding how to configure a product or service. For example, we are all aware of its power for helping an automobile manufacturer decide how to price a new model and determine the best name, shape, size and features.

TRADITIONAL VIEW OF CONJOINT ANALYSIS



At McKinsey, conjoint analysis is an important part of strategy development. By strategy development we mean a 3 step process:

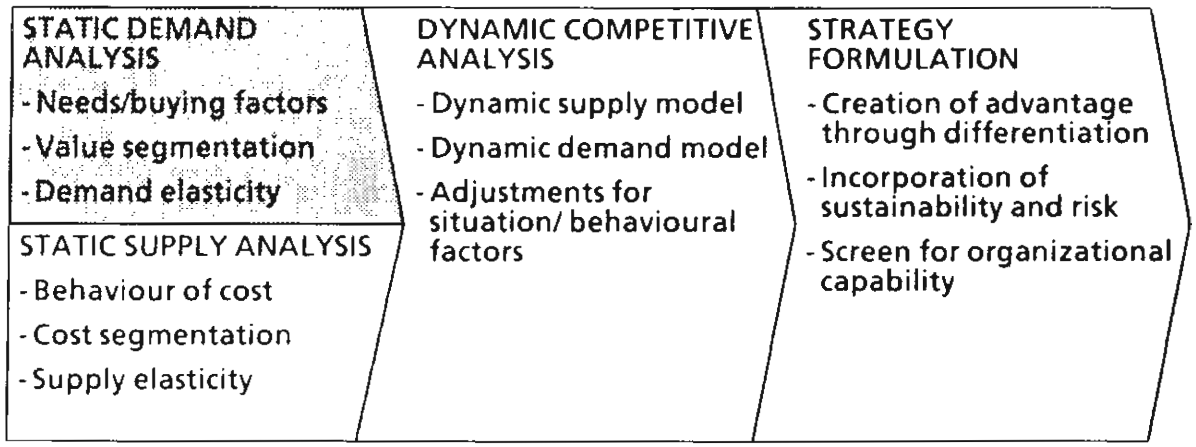
1. First of all we look at industry structure in terms of static supply and demand.

2. We then look at the dynamic picture. That is, how the industry structure is likely to change over time as a result of competitors' conduct.

3. Finally, based on insights gained from the static and dynamic situations, we formulate a strategy to achieve superior performance.

Conjoint analysis is the tool we often used for conducting demand analysis in Step 1. Several years ago the Firm evaluated alternatives ranging from econometric modelling to more basic market research techniques. We concluded that conjoint is the most powerful and appropriate technique. In addition, it offers the ability to provide insight into both the static and the dynamic demand issues.

OVERVIEW OF STRATEGY DEVELOPMENT MODEL



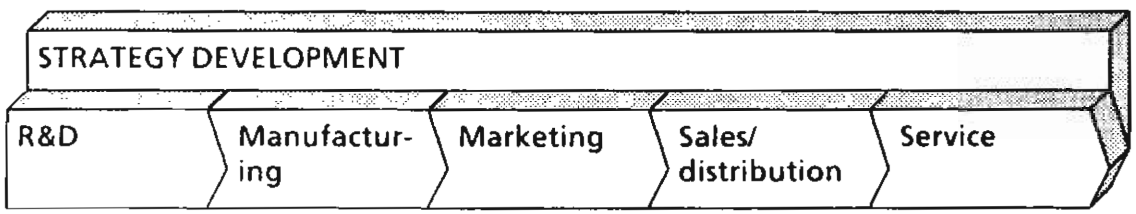
CONJOINT ANALYSIS: A STRATEGIC TOOL, NOT JUST A RESEARCH TECHNIQUE

- Applicable to any activity in a company from R&D to after sales service

- Applicable to corporate decisions on allocating resources between activities

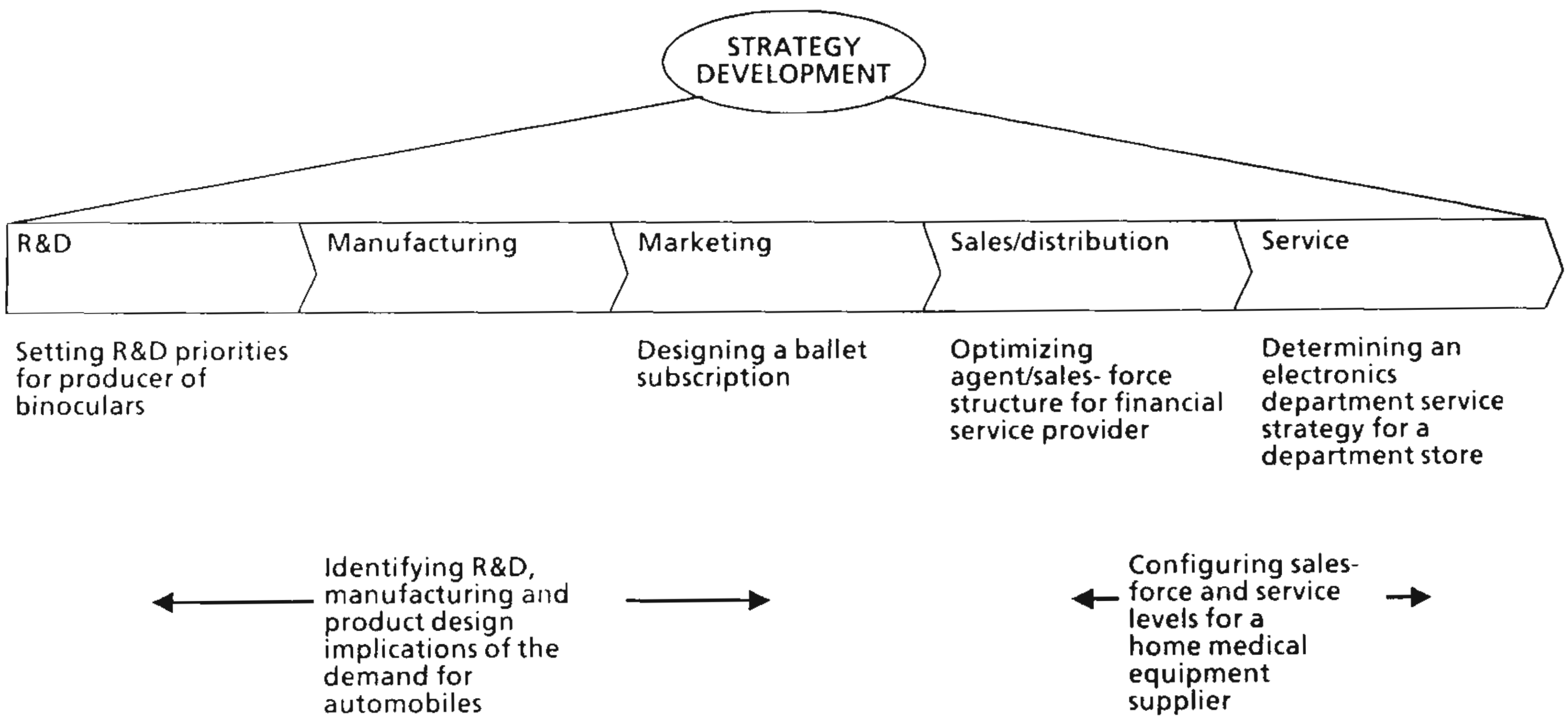
Strategy development is broader than just product features. It relates to each activity of a company, from R&D, to manufacturing, to marketing and through after sales service. There are opportunities to apply conjoint analysis at each stage of the business system.

BUSINESS SYSTEM APPROACT TO APPLYING CONJOINT ANALYSIS

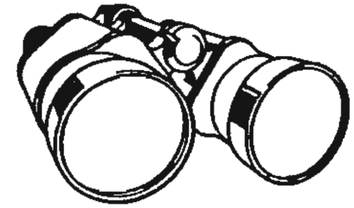


Some conjoint studies have addressed issues specific to one area of the business system. Others have explored two or more stages of the business system. In the following pages we describe three examples.

APPLICATIONS OF CONJOINT ANALYSIS ACROSS THE BUSINESS SYSTEM



Typically, to understand R&D priorities, one would examine manufacturing strengths/weaknesses and technological trends. In this situation, instead we first used conjoint analysis to determine what was important to the customer. We then explored the R&D implications of the important features.

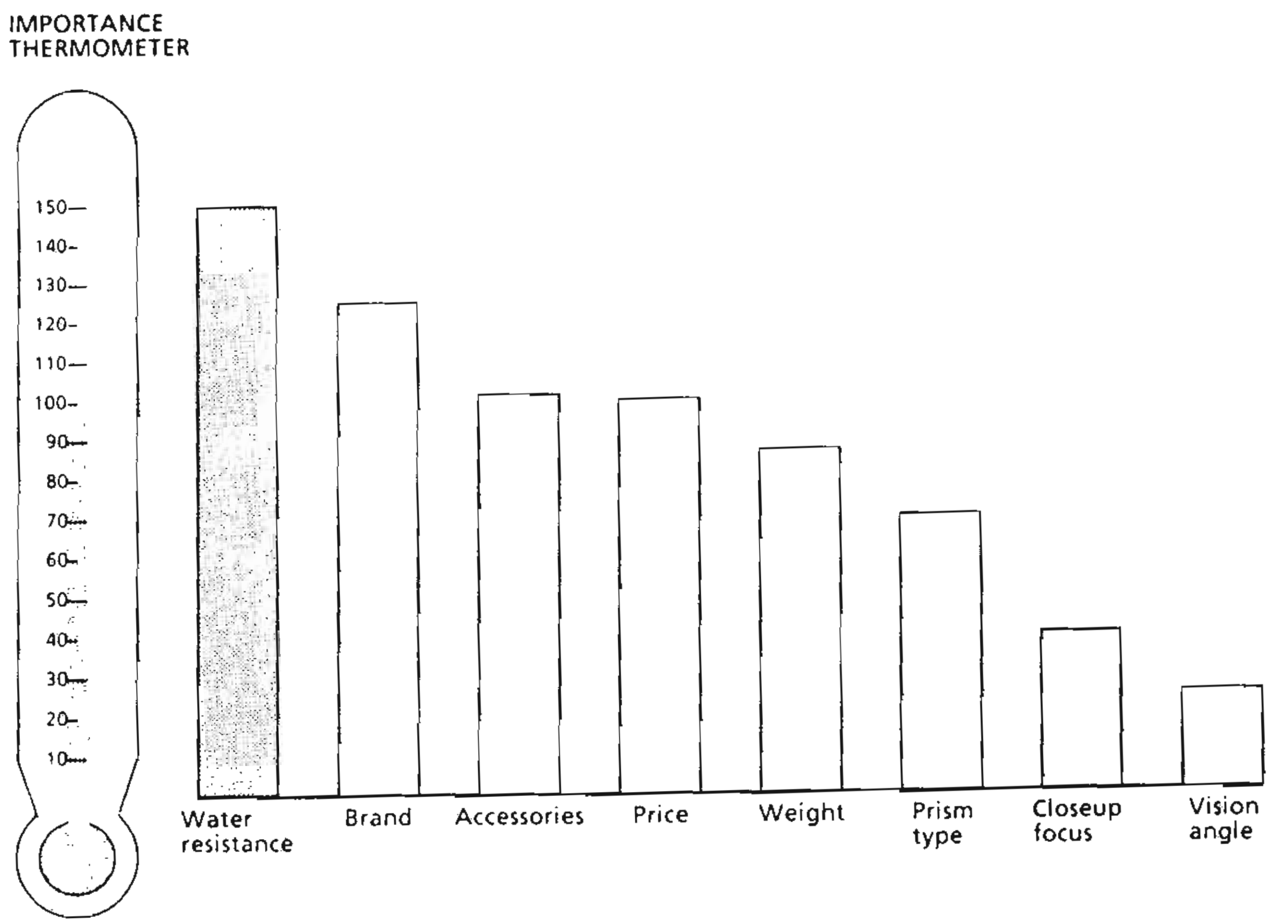


EXAMPLE 1: 'THE WRAPPER CAN BE CRITICAL'

Key questions answered by conjoint analysis	What are the drivers of buying behaviour in the high end market?
	What attributes should R&D focus on?
	What is the current value of the brand in this segment?
	Could the client get a premium price?

The client, an optical manufacturer, expected technical issues to dominate. Of great surprise was the amount of importance associated with water resistance. In fact, it was more important than all technical attributes combined. This was particularly true among hunters who represented the high end of the market.

RELATIVE IMPORTANCE OF ATTRIBUTES
price indexed to 100



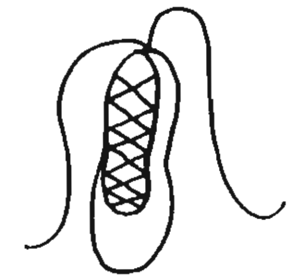
Source: Conjoint analysis

RECOMMENDATIONS

¶ R&D should broaden from only technical priorities to include water resistance.

¶ Raise prices 5-10 per cent.

In a pro bono project for a ballet company we were asked to help configure the subscription package. From interviews with the ballet's sponsors and patrons we learned there was strong support for keeping 'The Nutcracker' - it was thought to be an indispensible part of the community's Christmas experience. It was clear that to change this point of view would require a strong fact based approach. To address this issue without running the risk of losing financial support for the ballet we conducted a conjoint project.

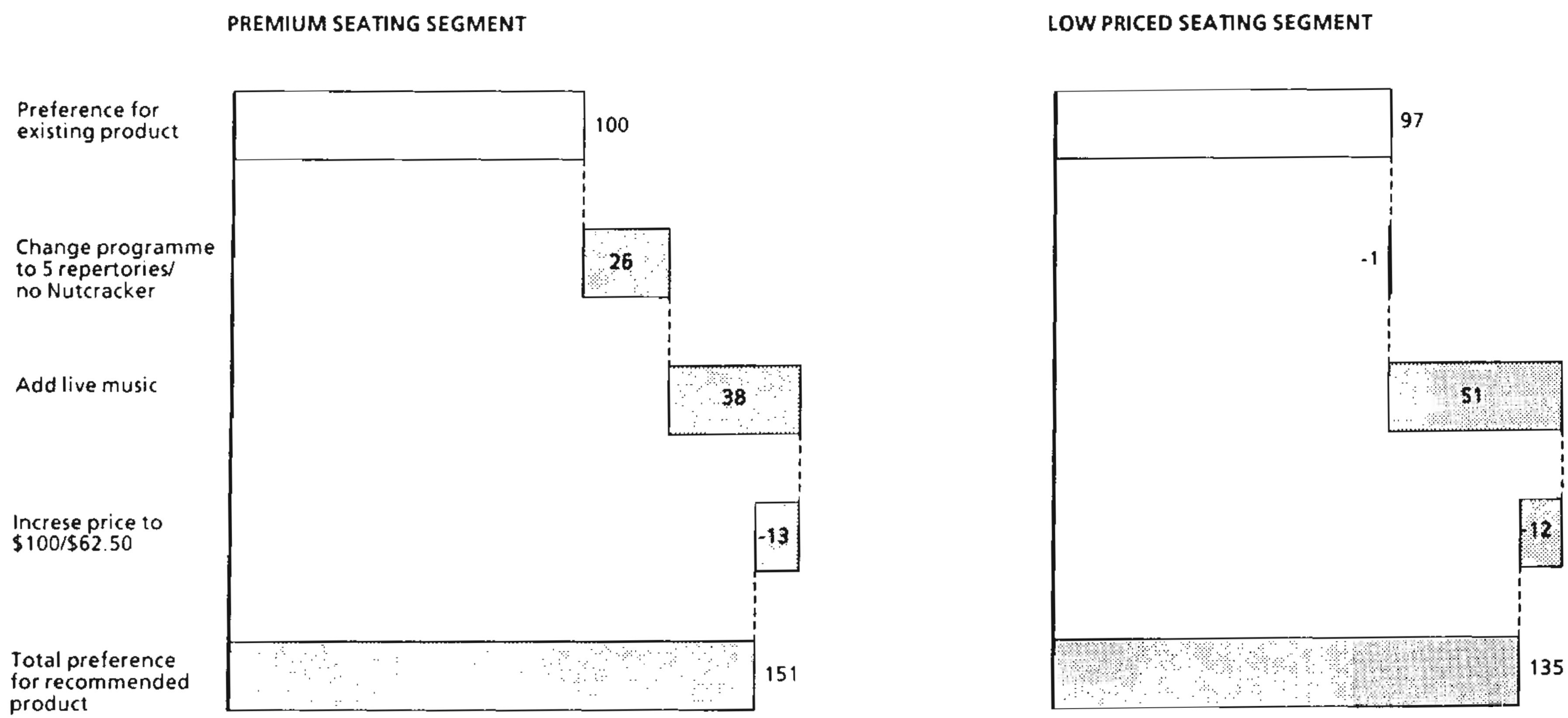


EXAMPLE 2: 'A HARD NUTCRACKER TO CRACK'

Key questions answered by conjoint analysis	What subscription package attributes are most important to ballet audiences? Could they profitably increase prices? Which combination of attributes would create the largest subscription demand?

The conjoint analysis showed there was little interest in maintaining the Nutcracker in the community at large. In light of these findings the patrons/sponsors were much more willing to consider changing the programme mix.

RELATIVE PREFERENCE - MODIFIED VS. EXISTING PROGRAMME
points illustrate incremental value



Source: Conjoint analysis

RECOMMENDATIONS

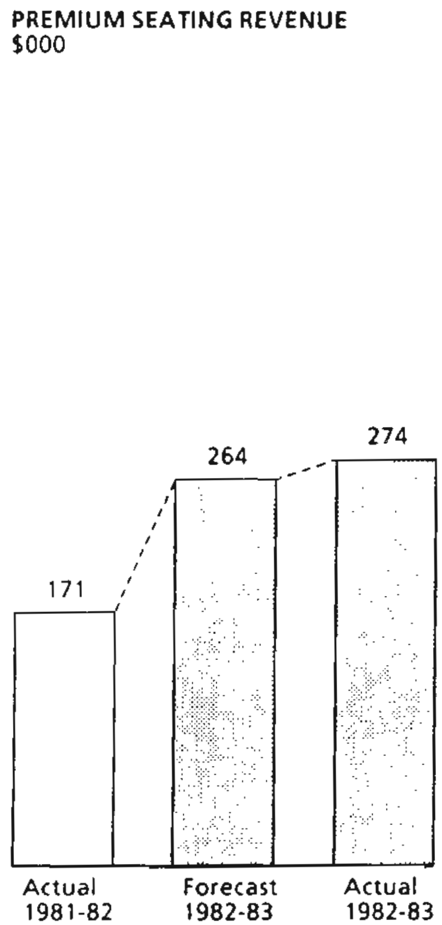
¶ Drop the yearly production of 'The Nutcracker'.

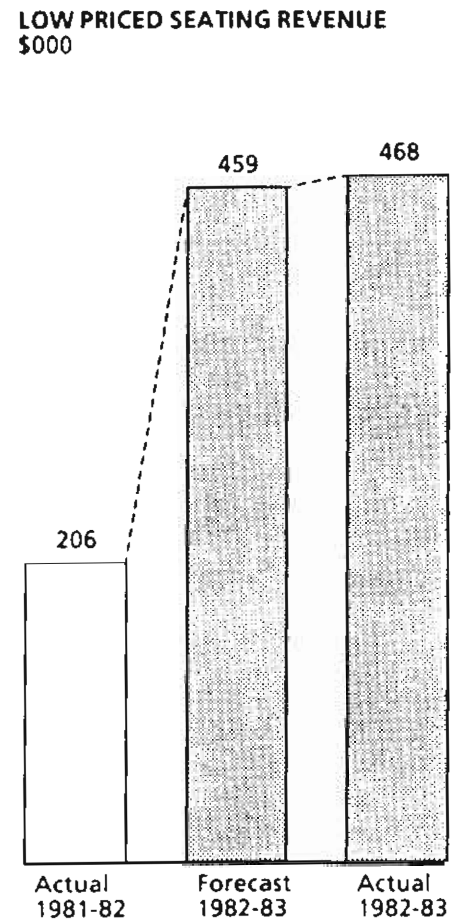
¶ Raise prices.

¶ Add live music (subject to costs).

A common question is how well conjoint analysis forecasts demand. Data are rarely available to address this issue, either due to a change during implementation or environmental conditions. In this case we were able to compare forecasts to actual revenue. The conjoint analysis provided an accurate forecast of the ballet's performance.

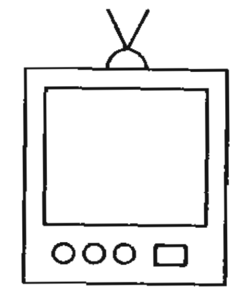
CONJOINT FORECAST VS. ACTUAL REVENUE





Source: Conjoint Simulations; team analysis

In an attempt to reduce overhead expenses most department store retailers are cutting back on after sales service. This is particularly true for delivery and repair services. The client asked us to help them identify areas where they could cut back on costs. A conjoint study, along with several other key analyses, provided somewhat unexpected answers.

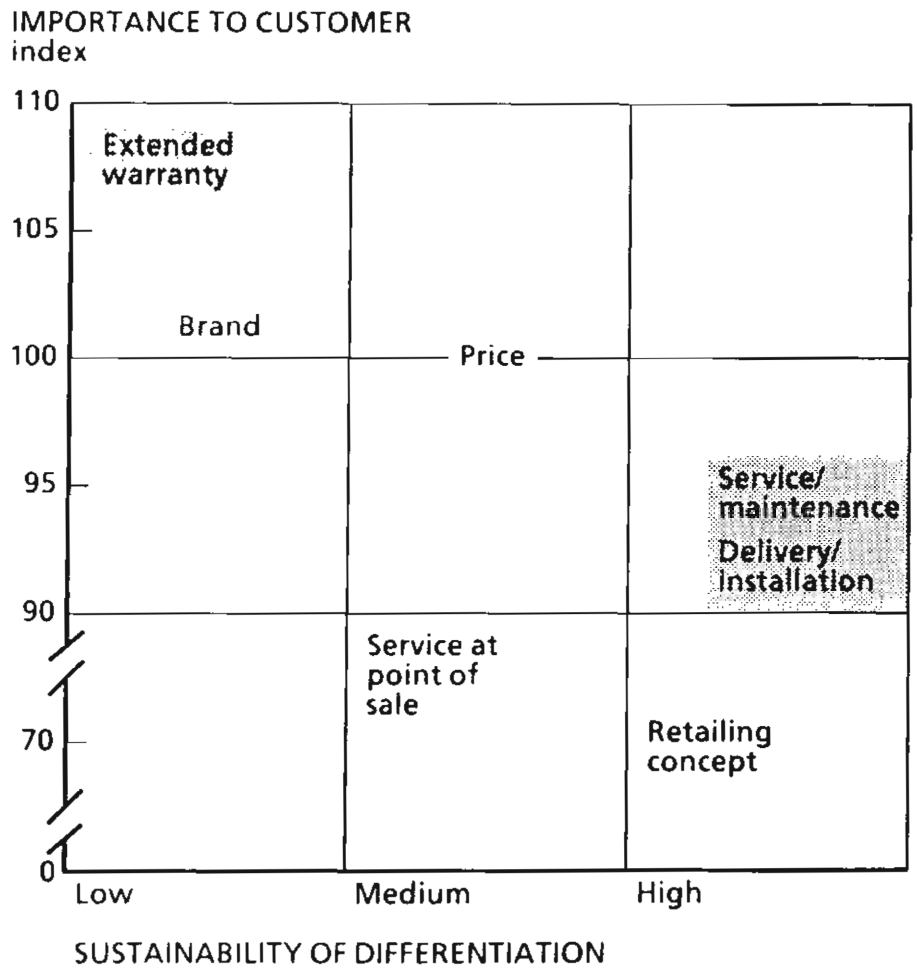


EXAMPLE 3: 'GOING AGAINST THE TIDE IN AFTER SALES SERVICE'

Questions answered by conjoint analysis	How important is an inhouse service organization to the customer?
	Could the client increase profits by offering extended warranty contracts?

After sales service, while not the most important buying factor, offered the greatest possibility of sustainable differentiation. Extended warranties proved to be important but competitors already offered this service.

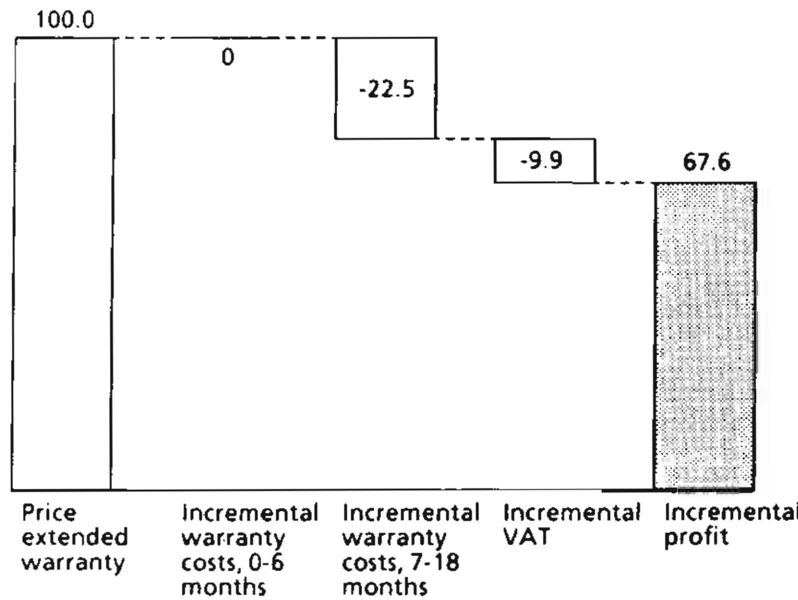
IMPORTANCE VS. SUSTAINABILITY MATRIX



Source: Team analysis

However, extended warranties proved to be very profitable. They increased the demand for the client's overall retailing concept and assured greater use of the service organization. In essence there was a significant synergy between after sales service and extended warranties.

PROFIT IMPACT OF EXTENDED WARRANTY CONTRACTS
$/unit



Source: McKinsey analysis

RECOMMENDATIONS

¶ Maintain delivery and installation service.

¶ Expand maintenance service organization.

¶ Offer extended warranty at a price premium.

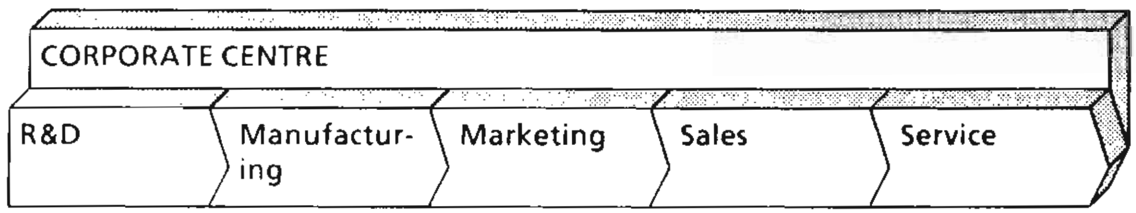
CONJOINT ANALYSIS: A STRATEGIC TOOL, NOT JUST A RESEARCH TECHNIQUE

Applicable to any activity in a company from R&D to after sales service

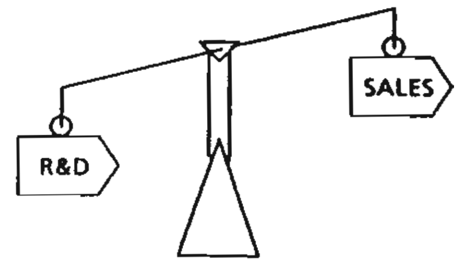
Applicable to corporate decisions on allocating resources between activities

Although we often look at the corporation as a series of separate activities, in practice, the corporate centre must set priorities and allocate resources among the parts of the business system. For example, should money be invested in R&D or advertising? Should salesforce be added or corporate staff increased?

CORPORATE ISSUES



The client, an industrial goods distributor, asked to us to help them quickly improve short term profits, or risk bankruptcy. Our recommendations would need to motivate a highly contentious group of vice presidents to act as a team and implement solutions.



CORPORATE STRATEGY

Key questions answered by conjoint analysis	What alternatives exist for improving profitability? Which of the alternatives are most likely to be implemented?

After identifying 12-15 options we helped the President narrow the alternatives. Based on the following six actions, a conjoint task was developed:

¶ Improve sales skills and incentives.

¶ Add sales people.

¶ Add new lines.

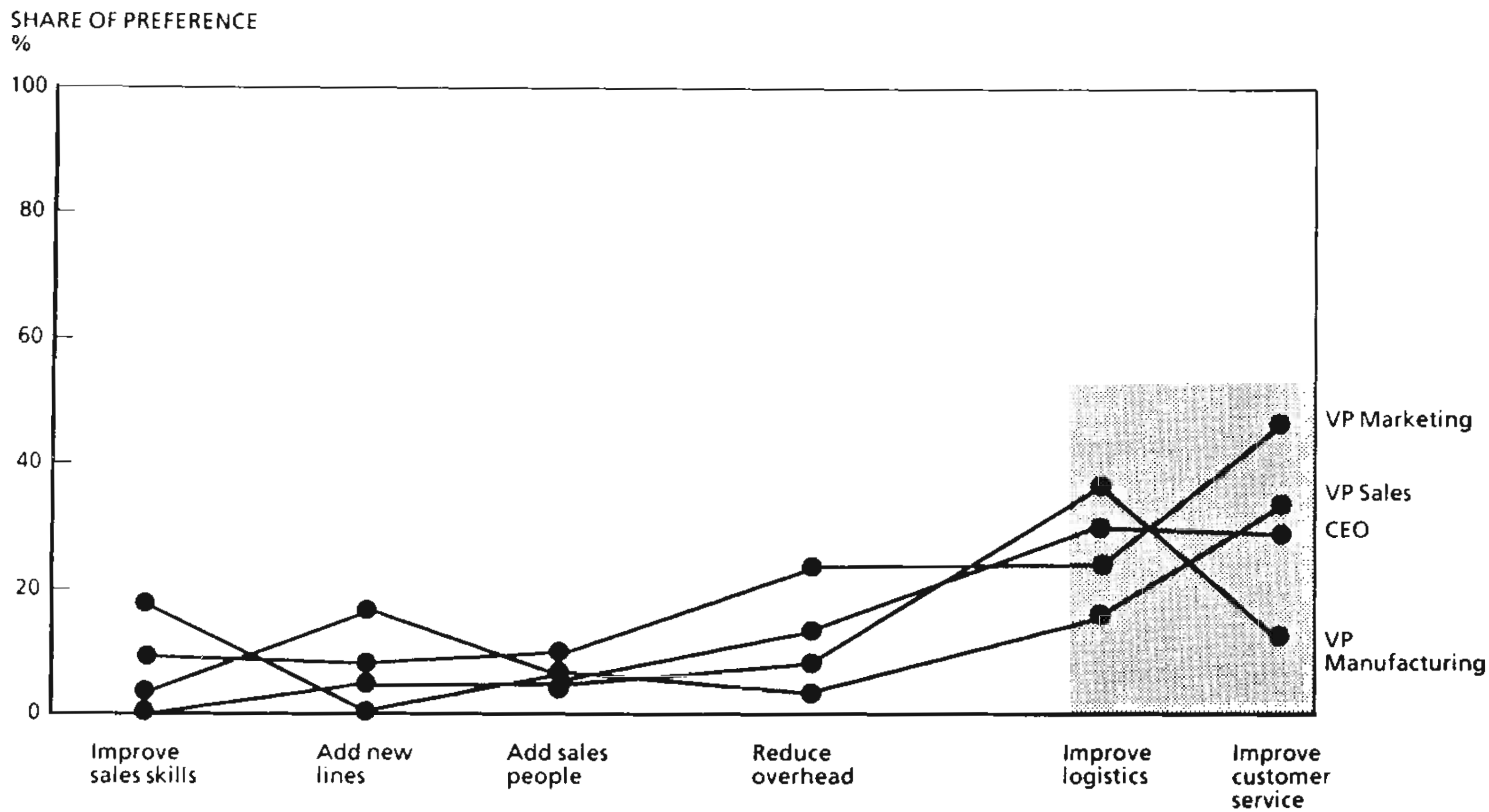
¶ Reduce overhead.

¶ Improve logistics (purchasing, inventory and warehousing).

¶ Improve customer service.

The findings showed that everyone could rally around improved logistics and customer service as a way to improve short term profitability. Orders shipped complete and on time improved from 65 per cent to over 95 per cent.

EXECUTIVE RATINGS



Source: Conjoint analysis

LESSONS LEARNED

Based on McKinsey's involvement in over 200 conjoint projects worldwide we have learned the following:

¶ Conjoint analysis is an extremely flexible and versatile tool that can be applied to any business decision where allocating resources is an issue.

¶ Particularly for manufacturers, conjoint analysis with end users is useful for broadening the client's understanding beyond technical concerns.

¶ Conjoint analysis is an extremely powerful tool to get views of individuals, not just groups of consumers.

¶ It can help to break down myths/misconceptions and radically change points of view.

LESSONS LEARNED

Flexible, versatile tool

Extends beyond technical concerns

Gets individuals' views

Can change points of view

USING CONJOINT STRATEGICALLY TO ENHANCE BUSINESS ENGINEERING

Roger Moore
The Boston Consulting Group

This paper examines two ways in which conjoint analysis can help a company strategically. First, the conjoint process can make an organization more aware of the business environment in which it competes. Second, conjoint analysis can allow the company to engineer/re-engineer a business to compete effectively in the marketplace. The Boston Consulting Group (BCG) has been using conjoint for a decade to help its clients develop business and product strategies. We have had the opportunity to refine the business engineering processes in which conjoint is a key tool. We have also been able to "post audit" the predictions and to track "plan versus actual" over strategic time frames.

This paper first discusses how conjoint analysis can lead to strategic advantage and then concludes with an example of how conjoint was used successfully to engineer a business.

To lead to strategic advantage, a conjoint analysis must focus on business-space attributes, as opposed to product-feature or product-space attributes. Business-space attributes are those characteristics of the company's overall business that affect the customer's buying decision. These attributes include things such as brand name, service reputation, and product availability/distribution.

TABLE 1

THREE LEVELS OF SPECIFICITY

Product Feature	Product Space	Business Space
Doors	MPG	Brand Loyalty
Cargo Space	Acceleration	Service
4WD	Ride	Parts Availability
Antilock Brakes	Size	Resale Value
Towing	Price	Reliability
Seating	Functionality	Product Performance

On the other hand, product-feature attributes include "engineerable" or tangible product features that can be determined by examining the product, such as number of doors or the amount of cargo space. A conjoint conducted at this level is used mainly for product development.

Product-space attributes are of two types: aggregates and product-related attributes. The aggregates include interrelated features that are too complex to understand individually, but that are easily identified when aggregated. These attributes tend to be functional in nature, such as "fuel economy." The product-related features would include such things as price and warranty. Traditionally, marketing conjoints are conducted at this level.

Business-space attributes are of strategic importance because advantages achieved in these areas can be defended over time more easily than product-feature based advantages.

Historically, conjoint analyses have more often focused on product-feature and product-space attributes than on business-space attributes. The infrequent use of business-space attributes in conjoint is typically caused by a directive from management to market research to understand the product and not the business. Management often does not realize the extent to which a conjoint can be used. If instead of just market research, a small task force including individuals from all disciplines (management, marketing, sales, and engineering) worked together developing the conjoint, the set of attributes would evolve into a set more relevant to the business space.

BUILDING AWARENESS

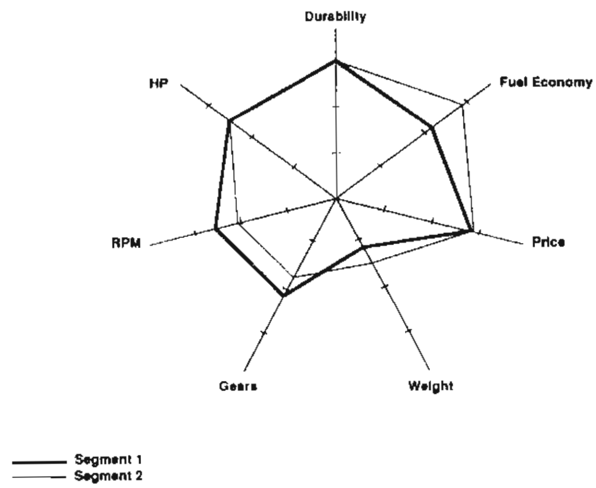
The advantages that stem from the results of a conjoint conducted for business engineering are well known. However, less well known are the advantages that stem just from conducting the process. The conjoint process can be very valuable in helping an organization become more aware of and sensitive to its environment: its customers' needs and the position of its current or planned products in the marketplace. In many cases, the exercise of defining the business-space attributes and their levels is more important to the organization than the results of the conjoint.

The conjoint focuses an organization's attention on the marketplace, because the first step in conducting a conjoint is to determine the attributes that are most important to customers. Typically, the first attempt at defining the attributes generates attributes more important to the organization than to the customer's buying decision. However, the conjoint process should include several iterations which involve both meetings among different parts of the organization and field interviews that test attributes and levels. As the organization works through this process, it begins to take on the customer's perspective and improves its ability to think strategically about the marketplace. The company can therefore make more informed decisions about product changes or offerings.

Once the organization understands the customer's perspective and has laid out the attributes key to the customer's buying decision, it can start thinking about how to engineer its business to establish a competitive advantage or niche in the business space. The attributes and levels developed from the conjoint process define the business space in which the company is working. Each competitor's product covers a part of the total business space; this defines its specific business approach. A helpful tool for visualizing the business space is the spider web diagram.

CHART 1

SEGMENT ENGINE PREFERENCES



Each leg of the spider web is an attribute and the levels of each attribute are laid out along the leg, with the least preferred being closest to the center and the most preferred being farthest from the center. Each competitor can be arrayed on the spider diagram to identify its business approach. The spider diagram, with all competitors arrayed, can help the organization understand the customers that each competitor is trying to reach and can increase the company's general awareness of the competitive environment. Along with identifying competitors' positions, the spider web diagram can also identify undeveloped business opportunities. It is very hard to offer a product that covers the entire business space--fundamental trade-offs exist between such items as complexity and cost. Finding an attribute or several attributes whose preferred levels have not been covered by another product's business approach identifies a possible business opportunity for a new product.

BUSINESS ENGINEERING

Once the attributes and levels have been established, the data can be collected and analyzed to begin the business engineering exercise.

One analysis that is particularly important to business engineering is determining market segments. The business engineering exercise is based mainly on simulations performed to determine optimum product offerings, and these offerings are likely to vary by segment. Consequently, the simulations should be based both on an understanding of the business space developed during the initial phase of the conjoint and on the segments identified during the analysis. The spider diagram may indicate an unserved market that a single product cannot cover due to fundamental trade-offs of the attributes. Then the segment information can identify whether none, one, or all of the markets actually have customers willing to purchase the products indicated by the business space.

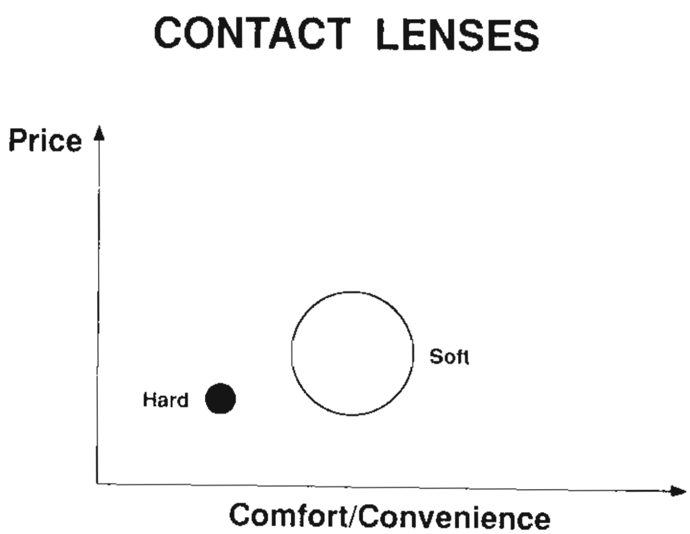
BUSINESS ENGINEERING EXAMPLE

The example below depicts how conjoint was used at Cooper Labs to engineer the extended wear contact lens business. Cooper Labs was a manufacturer of eye solutions prior to the introduction of the extended wear contact lens in 1980.

The contact lens market developed in the 1960s with the introduction of hard lenses. They had no patent protection and did not require FDA approval. The lenses were manufactured by local unbranded labs on manual lathes.

In the 1970s Bausch and Lomb introduced soft contact lenses. The new lenses were patent protected and required FDA approval. The lenses were manufactured with a special spin casting process. Bausch and Lomb obtained a 65% market share with the new product.

CHART 2



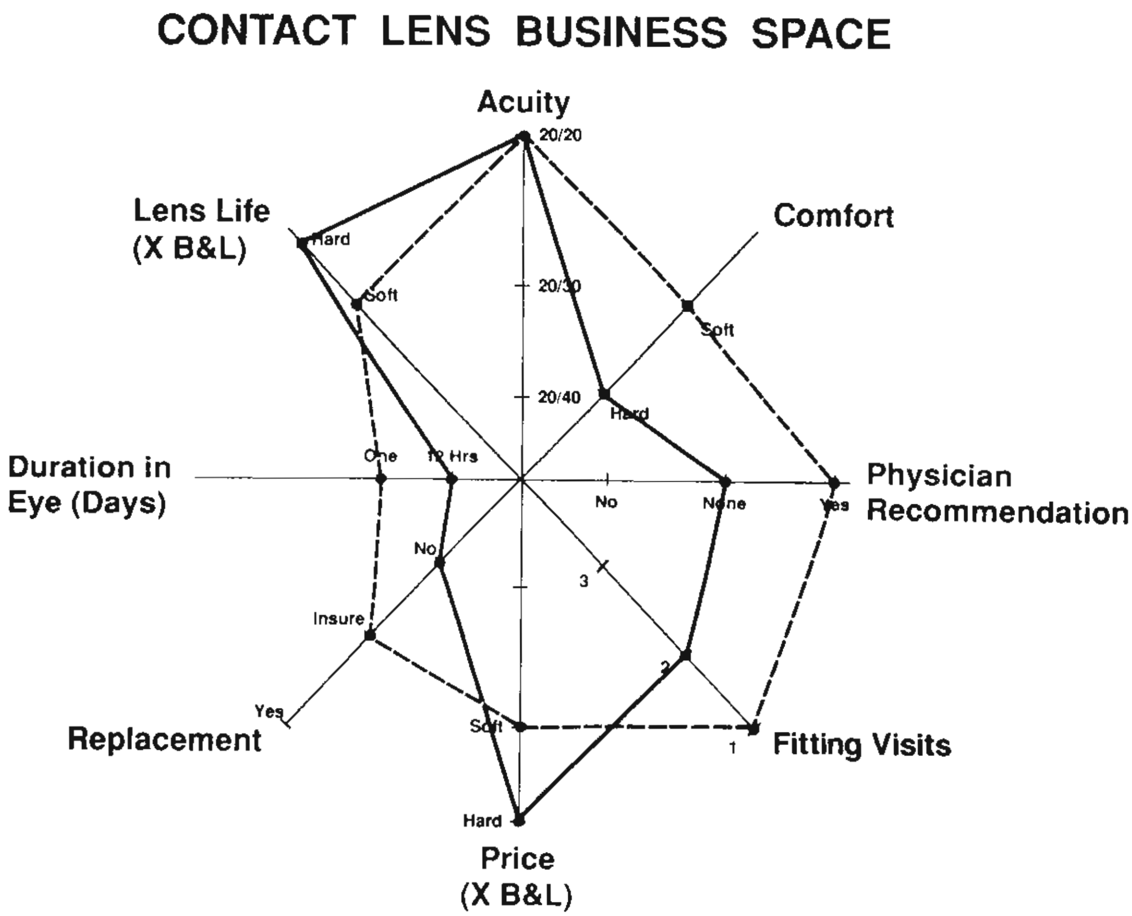
In the late '70s a new plastic polymer was developed in European labs. It was mostly water (similar to Jell-o), gas permeable, and organically inert. The new polymer was used as a bandage on the eye after cataract surgery. An attempt was made in Europe to use them for contact lenses, but the lenses provided poor visual acuity and would often just fall apart. They achieved at best a 6% share of the European market before being removed.

In 1979 Parker Montgomery, CEO of Cooper Labs, was on a ski vacation in Europe and heard about the new polymer; he was excited about the possibilities and believed that the technical problems could be overcome. Montgomery purchased rights to the new polymer and took the idea back to the organization. Montgomery received less than an enthusiastic response. Marketing felt that it would be a high cost, low volume item; the legal department was concerned at the amount of trouble required to get FDA approval for the new product; manufacturing believed the untested precision molding technology required to make the lenses could not be easily implemented; and the corporate planners asserted that taking on Bausch and Lomb with its 65% share was not a sound move.

Undaunted, Montgomery came to The Boston Consulting Group to find a way to convince the organization that the new extended wear lenses would succeed. After working with Cooper Labs it was evident that the legal, manufacturing, and planning issues could be resolved; however, it was unclear if marketing was correct about the low potential-the lens had been tried in Europe and failed. To resolve this issue and help specify key business parameters, the company needed to understand the contact lens market and consumer requirements. BCG proposed a conjoint analysis to determine if the new lenses were feasible.

Work with the client organization, doctors, and potential users developed the attributes and levels suggested by the spider diagram below. On this spider web of the business space, the clear superiority of Bausch and Lomb is evidenced by the larger area of its business space, which includes most of the hard lens profile.

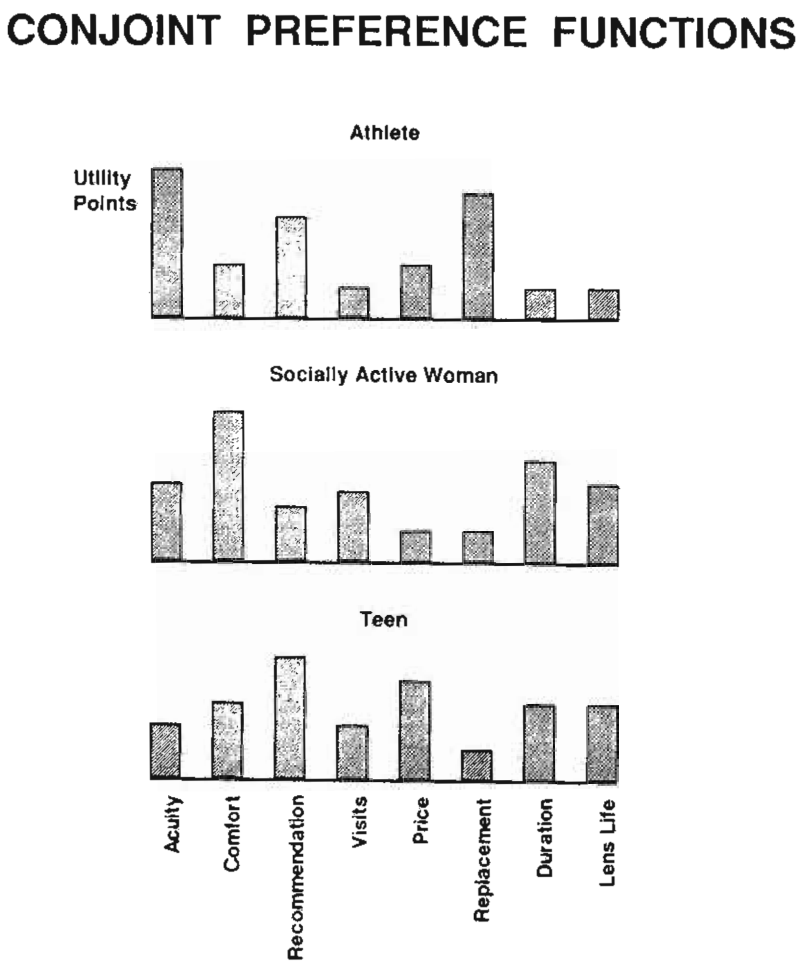
CHART 3



Soft lenses offered advantages over the hard lenses on almost all attributes. However, the spider diagram also clearly indicated room in the business space for a new product.

The conjoint data was collected using trade-off matrices (this was before computerized conjoint interviewing software was available) on a large, stratified sample of current and potential contact lens users. Early analyses indicated clearly explainable segmentation. Examples of three segments are given below.

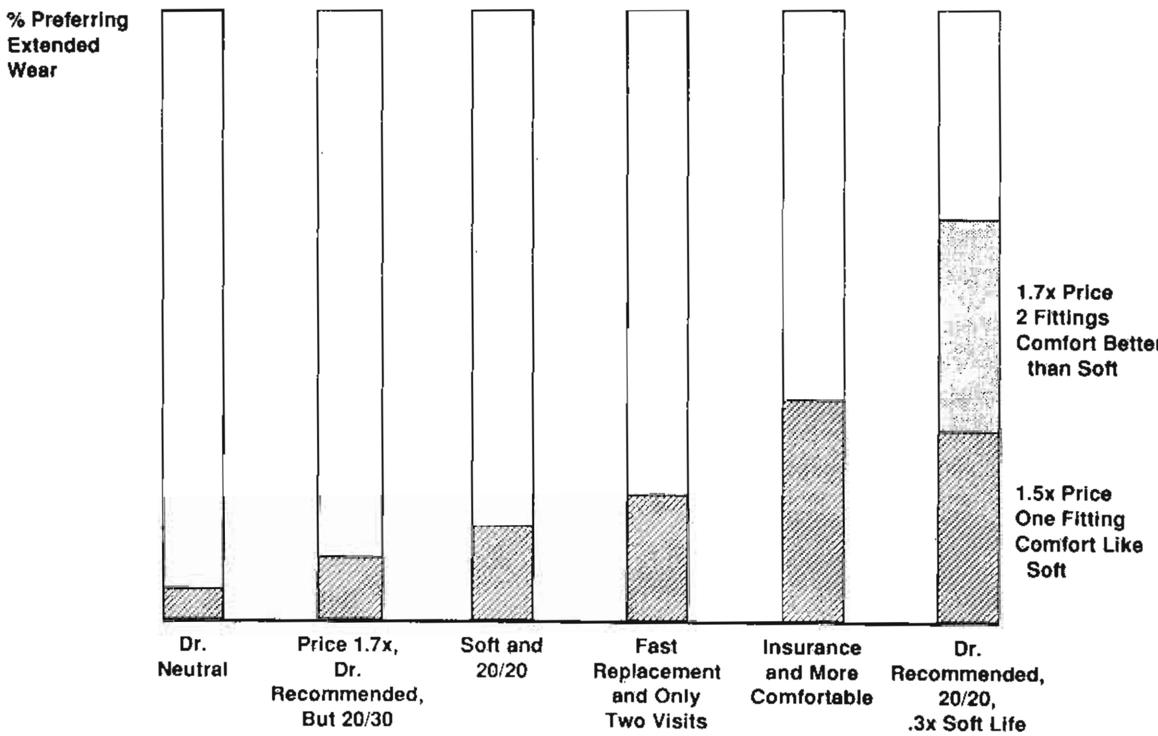
CHART 4



Using a market share simulator, possible products were developed and their market share potential evaluated.

CHART 5

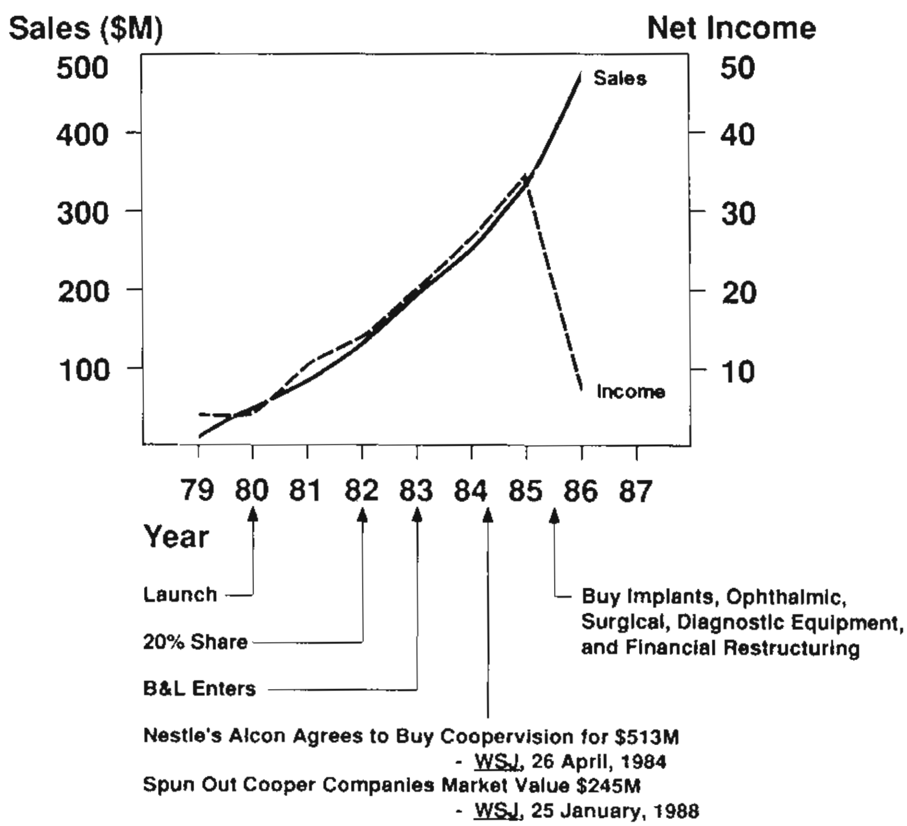
BUSINESS ENGINEERING WITH CONJOINT



Simulating the product that had been offered in Europe showed that the U.S. market would be similar; at best the extended wear lenses would achieve a very low share. But as business and product features were enhanced, it was clear that share could be increased. However, improving the attributes that could be manipulated with business positioning and technological innovation, the best a single product could do was about 30% market share. The segmentation discovered in early analyses suggested that more than one product was needed to satisfy the market. Simulations showed that two related, but differently positioned products could achieve close to a 60% market share of all potential wearers. In 1980 Cooper launched their two products, Permalens and Permaflex.

CHART 6

COOPERVISION PERMALENS AND PERMAFLEX



By 1982 Cooper had achieved a 20% market share. Bausch and Lomb was not able to respond with products of its own until 1983. In 1984 Nestle's Alcon agreed to buy Coopervision for $513 million. Bausch and Lomb filed an antitrust suit with the FTC and blocked the purchase in 1985. At this point the story takes a turn for the worse. Left with a half-million dollar business and undervalued stock, Montgomery feared corporate raiders would try to take over the company. He used all available cash and borrowings to buy new medical businesses in order to grow, diversify, and ward off takeover attempts.

Cooper went through many financial and legal restructurings as part of this defensive process. The core lens business emerged as Cooper Companies with a market value of about $500 million before October 1987. In 1988 it had fallen to $245 million. Late in 1988 the company was, indeed, taken over by raiders just as Montgomery had feared.

CONCLUSION

As the Cooper example demonstrates, conjoint analysis can be used very effectively to engineer a business. At Cooper, the conjoint convinced a skeptical organization and helped develop a $500 million business in six years. However, as this paper points out, the conjoint process itself can provide even more important strategic benefits by heightening a company's awareness of both its customers' needs and its competitive environment.

GAINING A COMPETITIVE ADVANTAGE BY COMBINING PERCEPTUAL MAPPING AND CONJOINT ANALYSIS

Harla L. Hutchinson
John Morton Company

These proceedings outline a variety of interesting and useful applications for the techniques of conjoint analysis and perceptual mapping. I'll discuss how combining the two techniques can in some cases provide more information than can their separate use.

First, I will offer a fairly simplistic look at the two techniques, focusing on their typical uses and what I see as frequent shortcomings. Next I will present a situation we encountered with a client in which the two techniques were combined to provide a useful, and perhaps unique, perspective on a very complex market. I will conclude with a discussion of the general characteristics of a market that could benefit from this approach.

TYPICAL USES OF THESE TECHNIQUES

Both of these techniques have proven useful in a wide variety of market applications. Each provides a different kind of information and each is typically used for different purposes.

Perceptual mapping, for instance, is good at providing an understanding of what it is that distinguishes among the competitors in a given market. If all competitors were perceived to have the same general characteristics and to provide the same benefits, a buyer's choice among them would be very difficult indeed. In a situation like that, you might expect buyers to make selections randomly, and the shares of all competitors would be approximately equal (assuming, of course, equal availability and awareness). But since those of us out here in the real world don't expect to encounter that kind of situation, an understanding of the ways in which competitors are seen to differ can be very useful.

Once you know how buyers are distinguishing among products, the next logical step is to ask, "How do my market entries fare with respect to those distinguishing characteristics? How about my competitors?" Perceptual mapping can provide this kind of information as well, visually displaying the strengths and weaknesses of each competitor for all to see. This leads logically into an investigation of perceived similarities and differences among the competitors. From this you will frequently gain a better perspective on likely patterns of competition existing in the market -- competitors which appear near each other on a perceptual map are more likely to compete closely with each other than those who are relatively far apart.

While perceptual maps provide a great deal of information about the current structure of a market, seldom would you want to stop there in drawing your conclusions. The next step you would probably take would be to examine the map in an attempt to discover directions for change or areas of possible opportunity. In other words, what does it appear should be done to differentiate your market entries from their competition? Does there seem to be a spot for a single competitor which combines two different sets of benefits currently being provided separately by two different sets of competitors? As you can see, perceptual mapping can provide a great deal of useful information about a market.

Conjoint analysis is typically applied to a very different set of purposes. While perceptual mapping focuses on the characteristics which differentiate competitors, conjoint concentrates on the characteristics which influence preference. What factors are important to buyers in this market? How important is each relative to other considerations? What are the trade-offs the market is willing to make among the characteristics?

As with perceptual mapping, there is a great deal of usefulness in understanding this relatively static aspect of the market. But, there is more to be gained by using this knowledge to help direct future actions in a market. When used to simulate a market, conjoint analysis can do an excellent job of allowing one to test alternative market scenarios for one's own market entries as well as those of one's competitors. Demand for or acceptance of new concepts can also be assessed using these models. Looking at who wins and who loses when market changes are simulated can also provide some understanding of the patterns of competition at work in the market.

Conjoint analysis is most typically applied to relatively hard product attributes, but the technique is widely useful. It can and should be considered in any market situation in which customers are required to make trade-offs among features and benefits, where the "perfect" product or service (from the customer's point of view) cannot be obtained.

Both conjoint analysis and perceptual mapping can be used in a wide variety of situations to provide valuable answers to a lot of important questions. Sometimes, however, you will find yourself still lacking some key pieces of information.

Mapping, as I said, is frequently used to develop hypotheses about market gaps and promising directions for change. However, by itself, it does not provide a means to actually test those hypotheses. Sure, it looks like northeast is the way to go on your map, but how far should you try to move and how should you get there? Most importantly, what is your share likely to be at the desired destination?

Conjoint analysis provides such answers easily, but in many cases is inadequate to provide its user with a clear picture of the competitive structure of the market. What is it that distinguishes the competitors from each other? Granted, if the market has relatively few competitors and you are dealing with a fairly limited set of characteristics, the strengths and weaknesses of each should be pretty easy to spot. But what if you are dealing with a market in which there are as many as 150 to 200 competitors and 40 or so attributes are required to adequately describe the relevant buying factors?

Taking this somewhat simplified view of the world -- and I have admittedly stacked the deck somewhat -- it seems as if the two techniques provide a good fit for each other. What I'd like to discuss next is how we combined the use of these two techniques for a client and what kind of information the approach provided for him that hadn't been available before.

A CASE STUDY

One of our clients, a major manufacturer of passenger vehicles, desired some additional input into their strategic planning efforts. They came to us with the following broad objectives:

- to provide a means for segmenting the U.S. passenger vehicle market using consumer-based criteria
- to provide a capability for identifying the position of any vehicle relative to its likely competition
- to provide a means for testing for improved acceptance of any vehicle through actual physical changes or changes in positioning, and
- to provide a capability for identifying and assessing the potential for new vehicle opportunities.

Since the first two objectives could be met through the use of perceptual mapping and the second two through conjoint, it was decided to use both techniques. Further, it seemed that the greatest usefulness could be achieved if both techniques were applied to a common set of vehicle characteristics.

The result of this investigation was the construction of a simulation model of consumer preference and perception in the U.S. passenger vehicle market. The model is primarily based on consumer perceptions and permits strategies to be tested for their effects on both share and positioning. In addition, linking the two techniques permitted the client to develop a much better understanding of the competitive structure of the market.

And how was this outcome accomplished? Believe me, it wasn't easy. Our first foray into this area began in 1981 and we continue to refine our approach today. I'd like to tell you about the general approach we have taken and to point out some potential problem areas along the way.

As with any research study, the first we first had to set the scope of our efforts. First, we had to identify the market to be covered. Was the client interested in every new vehicle available for purchase in the U.S., regardless of its price, share, or availability to the general public? Perhaps instead, the focus should be on the those market segments in which the client currently competes or hopes to compete. Another concern in this market, which may not have parallels in many other markets, was what should be done about nearly identical vehicles offered under different nameplates (examples would be Ford Tempo and Mercury Topaz, or Dodge Aries and Plymouth Reliant, or all the "who can possibly keep those cars straight" similarities across the GM lines): Should both (or all) be included, or should the highest-volume line stand as proxy for the rest? And finally, what should be done about all the variants available of a single car line - the GX, the SE, the LE, and the who-knows-what else? Should perceptions be obtained separately for each, or for the car line as a whole?

In the end, these issues were decided by the client after a careful evaluation of his needs from the model, coupled with market knowledge and balanced by sample size implications. The first model we built focused on the car market. Later a similar model was built for the truck market. It soon became apparent that interaction between the two markets, especially with the introduction and subsequent popularity of minivans, required that the approach be expanded to encompass all passenger vehicles. Today's model includes nearly 200 of the best selling vehicles in the country, which together account for more than 90% of all new vehicle sales. The specific categories included range from luxury cars to econoboxes, from sports cars to sport utility vehicles, and from family sedans to pickup trucks.

The next step was to determine the attributes of importance and interest. We wanted to be sure we included not only those attributes likely to enter in the decision of almost every consumer (such as price, quality, reliability, and manufacturer) but also attributes likely to be much more important to some groups of buyers than to others (such as acceleration, luxuriousness, practicality, and styling). In this process, representatives of various client departments, such as product planning and corporate strategy, along with representatives from their advertising agency and John Morton Company, worked together to develop an extensive list of possible attributes. Early on, the decision was made to stick with product and image attributes and not to extend the list to include such "external" considerations as financing and rebates.

These early lists were evaluated both in a series of focus groups and quantitatively to yield the approximately 50 attributes used in building the model. The attributes are worded from a consumer's point of view rather than an engineer's (for instance, we deal with "front seat roominess" ranging from poor to excellent, rather than trying to actually quantify the number of inches of head or leg room available). A few attributes are included which deal with the most important "hard" characteristics of vehicles, like manufacturer, price, body style, and transmission. Remember, this was primarily intended to be a perceptual model of the market.

Since we hoped to be modeling behavior at the individual consumer level, we felt that before designing the simulation system we also needed to better understand the consumer buying process. With such a large set of competitors from which to choose, how does the consumer go about narrowing down his choices to a more manageable set? This issue was also explored in the focus groups.

Armed with the answers to all these questions, we were ready to jump in and get down to business. A scope as broad as I have just described required quite a large data collection effort. Using broad age and sex guidelines and some specific vehicle type quotas, interviews were obtained with 2100 recent vehicle purchasers in fifteen cities around the country. They were pre-recruited from vehicle registration lists to central locations, where each participated in a computer-administered personal interview.

Each respondent began by selecting a set of vehicles he would consider as possible replacements to his recently purchased vehicle. No restrictions were placed on either the number or pattern of his choices, which is to say that we permitted the individual who purchased a Yugo to tell us he would also consider a Lincoln Town Car or a Ford Bronco. Since segmentation of the market was a major objective, it was felt this would provide much useful input.

Following this vehicle selection task, the respondent was seated at a computer for the remainder of the interview. From his set of vehicle considerations, five vehicles were chosen and he was asked to provide perceptual ratings on all attributes for each of the five. This was followed by a conjoint task, which focused on his fifteen most important attributes, after first obtaining rank order and relative attribute importance information for all attributes. The interview concluded with a short pyschographic battery and several demographic questions. The entire interview averaged about two hours in length.

We now had all the raw data necessary to begin constructing the model. Analysis began with an investigation of stated vehicle, manufacturer, and segment substitution patterns. These results, together with the perceptual maps, provided the basis for a new industry segmentation scheme. Consumer utility values were then calculated. All of this information fed into the construction of the simulation model itself.

MODEL FEATURES AND CAPABILITIES

The model has a base case as its cornerstone, which describes today's vehicle market in terms of consumer perceptions where appropriate, supplemented by a few actual vehicle specifications, such as price and the availability and cost of transmissions and so on. The user tests market changes by describing new vehicle concepts, changes in actual specifications or perceptions of current vehicles, and vehicles withdrawn from the market. These scenarios of future market conditions may be evaluated for their effects on positioning through a "perceptual simulation" and also for their effects on share through a preference simulation.

It may not be readily apparent at this point why this combined approach was so advantageous. I think I can illustrate with a few simple examples:

Let's start by looking at one of the basic perceptual maps to obtain some understanding of the market's structure (see illustration labeled "Car Industry Vectors"). What is illustrated here are the first two discriminant dimensions describing the U.S. car market. The east-west dimension distinguishes between economical and expensive vehicles, while the north-south dimension differentiates between fun vehicles and more practical ones. (Together these two dimensions account for about 65% of the variation. We also make use of a third dimension, differentiating primarily on quality, which accounts for another ten percent or so.)

Once we overlay the positions of the actual vehicles, as in the illustration labeled "Car Industry Plots," broad vehicle clusters emerge and we can begin to get an idea about the competition one's vehicle is likely to encounter. (To make this illustration just a little more concrete, the Ford products have been labeled). You can begin to see the kind of perspective which emerges: the closeness of one's own models within each market segment, one's absence or relative lack of presence in potentially important areas, the closeness (or lack thereof) of "expected" competition, early diagnosis of potential product problems and the possibility of emerging segments. Insight into each of these areas is available through careful analysis of the maps and the specific product perceptions which produced them.

From this analysis come ideas for strategies to be tested in preference simulations and for positioning changes. For instance, if Ford wanted to distance Taurus and Sable more from each other, the maps could suggest which attributes might prove most effective, while the preference simulation could indicate which distancing strategy is likely to prove most profitable. Since this model is based on consumer perceptions, it is important to note that the identified strategy might not require any actual changes to the vehicle, but might be completely achievable through a changed focus in advertising.

New product concepts can be tested in a similar way: Drawing on engineering specifications, comparisons to similar existing vehicles, and competitive intelligence, new vehicles can be "entered" in the market, with the model providing information about both their positioning relative to current vehicles and their appeal among consumers.

GUIDELINES FOR APPLICATION

While the scope of this effort may appear somewhat overwhelming, I think I can fairly state that the model has provided our client with a wealth of valuable information over the years, and I believe the approach can be profitably applied in other industries and markets. In attempting to determine its applicability to your particular situation, try to answer the following questions:

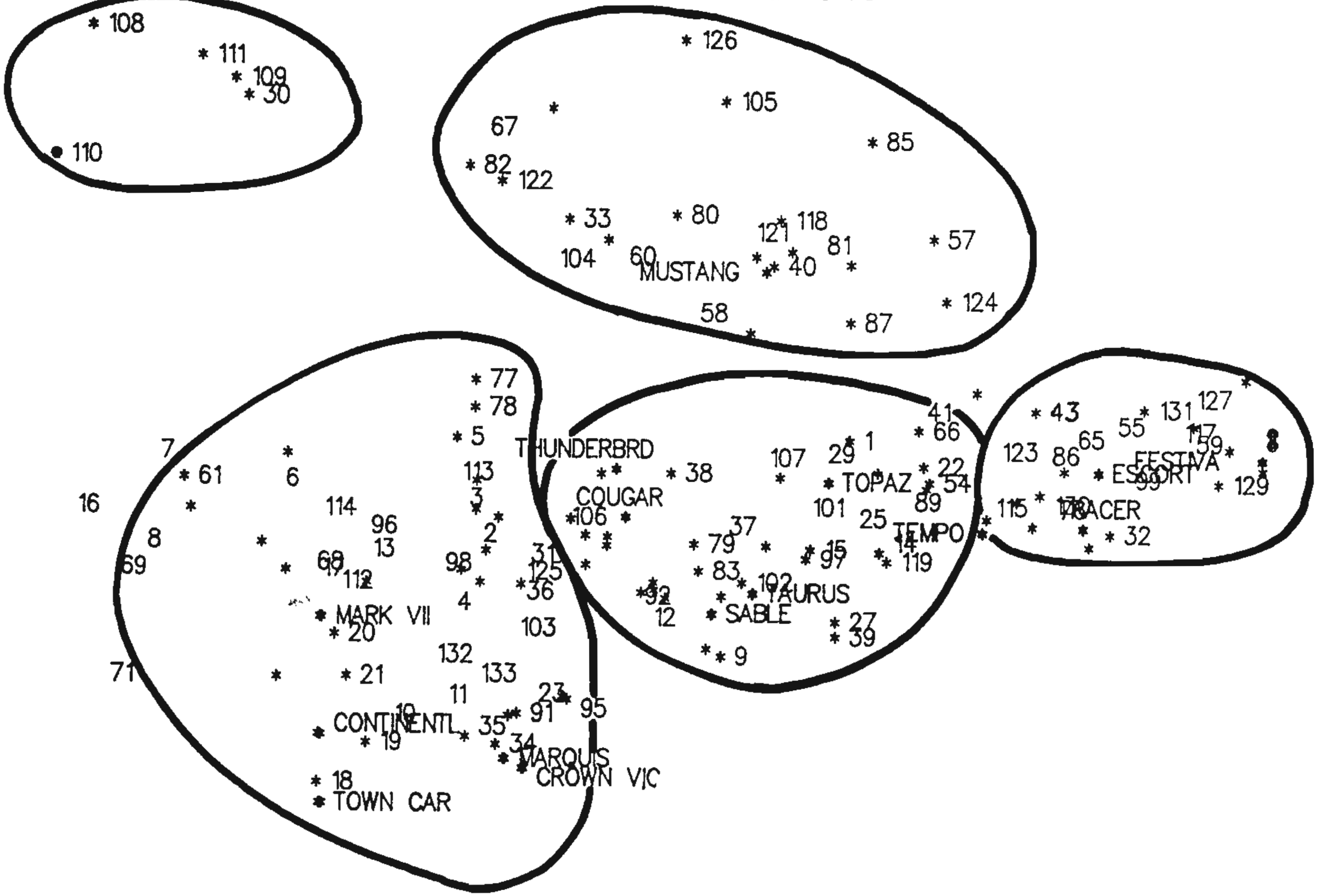
- Is the market driven more by image than concrete product or service characteristics? Do competitors differ with respect to how well they serve the needs of different consumers? If so, perceptual mapping is a natural technique to apply and an understanding of perceptions would be critical to the accurate specification of a conjoint model.

- Is the market large and complex enough to require considerable simplification to understand its dynamics? Smaller markets could also benefit from this approach, but the most advantage would be gained from its application in a more complicated market situation.

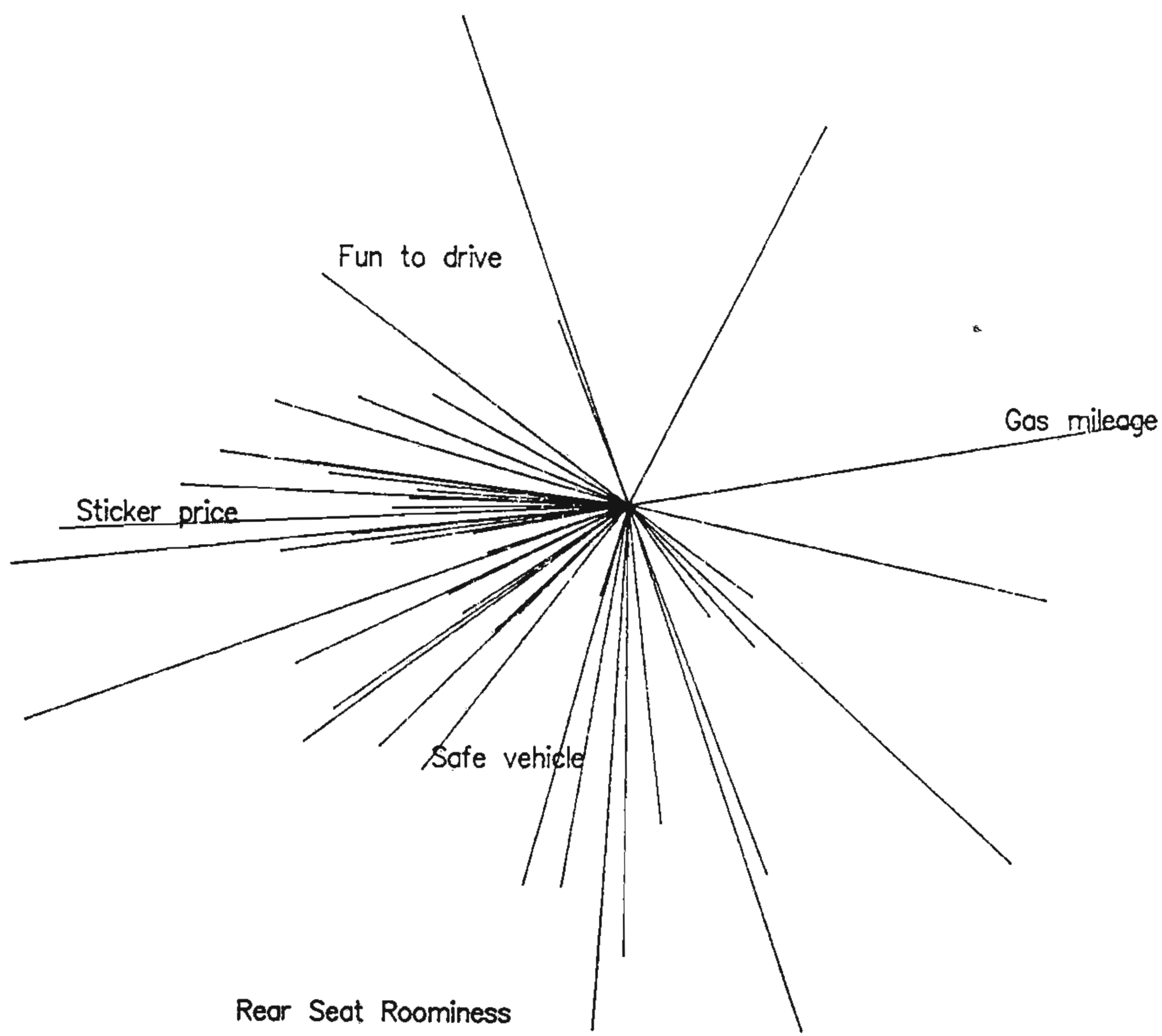
- Finally, can the key buying factors be identified? If questioned directly, can buyers tell you why they buy and what they value? Can the factors be worded in such a way to be appropriate for obtaining both utility values and perceptions? Keep in mind that you need to match utility levels to rating levels in some way, and also that you need to provide product planners with useful diagnostic information. This means, for instance, that treating an attribute about "ride" as something which ranges from firm to soft is going to be more useful than an attribute which deals broadly with "riding comfort." As a final caution in this respect, you should attempt to have as much overlap as possible between the two sets of attributes (though it may well not be realistic to obtain perceptions on every attribute or to word some image aspects in such a way that they may be traded off). This will avoid the situation in which certain strategies can be tested in one side of the model but not the other.

I hope these questions and my description of our approach in the automotive market have fueled your imaginations about possible applications to your own markets.

CAR INDUSTRY PLOTS



CAR INDUSTRY VECTORS



BIAS IN THE FIRST CHOICE RULE FOR PREDICTING SHARE

Terry Elrod and *S. Krishna Kumar*
Vanderbilt University

Abstract

Conjoint choice simulators let market researchers predict the share of potential customers who will choose a product that has been described in terms of its attributes. One of the simplest and most common methods for predicting share is the first choice rule. This paper presents an analytical investigation of the first choice rule when it is used to predict share for two alternatives. We show that under certain conditions the first choice rule yields substantially biased predictions of share. However, the first choice rule's two sources of bias, which work in opposite directions, can sometimes nearly cancel each other out, yielding nearly unbiased predictions of share. We also show when one commonly used rule for probabilistic prediction can reduce the bias and when it can only increase it. We expect our findings will generalize to predicting share among multiple brands.

Introduction

This paper is intended to help market researchers make better use of conjoint simulators for estimating share. By "share" we mean the fraction of potential buyers under study who would choose a concept. Our results generalize immediately to the case in which respondents are weighted by expected purchase volume or by their degree of representativeness of the population (Green and Kedia 1983), as long as the weights are uncorrelated with brand preferences.

We consider the case in which the researcher has administered conjoint analysis to a sample of potential buyers and has estimated attribute part worths at the individual level. This is a common method for administering conjoint analysis and has been found to work at least as well as analyses at the segment or market level (Wittink and Montgomery 1979; Moore 1980; Wiley and Low 1983; Elrod, Louviere and Krishna Kumar 1989). Once individual-level part worths have been estimated, share for any combination of attributes can be predicted using a choice simulator. Typically, the researcher must choose among several methods for choice simulation that do not yield the same share predictions.

One of the simplest methods for predicting share is the first choice rule. This rule predicts each respondent's utility for each brand in the choice set, and then assumes the respondent will buy the brand having the highest predicted utility. Its prediction of share is the fraction of respondents who are predicted to choose the brand. The first choice rule is the method, or one of several methods, for share prediction in many conjoint simulations (Green, DeSarbo and Kedia 1980; Urban and Hauser 1980; Cattin and Wittink 1982; Finkbeiner 1988). It is also used in several algorithms for optimal new product identification (Albers 1979; Green, Carroll and Goldberg 1981; Gavish, Horsky and Srikanth 1983; Sudharshan, May and Shocker 1987).

At first glance, the first choice rule seems to make perfect sense. But it assumes that respondents will always choose that brand having the highest estimated utility. This assumption is wrong in two respects (Louviere 1988). First, the conjoint attributes and part worths do not fully account for the utilities governing choice behavior. It is well known that consumers do not always choose their most preferred brand (Pessemier, Burger, Teach and Tigert 1971; Bass, Pessemier and Lehmann 1972; Reibstein 1978; Silk and Urban 1978; Morrison 1979; McAlister and Pessemier 1982). The discrepancies between the utilities governing choice and the conjoint predicted utilities will be larger the more the choice task and choice alternatives differ from the conjoint task and profiles. Yet even when respondents are asked to choose among conjoint profiles at the end of the conjoint task, prediction will not be perfect, in part because respondents sometimes make mistakes in their choice process. Some error term will always be required to capture the discrepancies, and yet the first choice rule assumes that no error term is needed.

Second, the first choice rule also ignores the fact that respondent part worths are estimated with error. It assumes that the brand with the highest predicted utility is the brand with the highest actual utility. But since the part worths are estimated with error, the brand with the highest estimated utility will not always be the brand with the highest actual utility. The probability of an incorrect prediction is greatest when the actual utilities are nearly equal.

Both objections to the first choice rule imply that choice predictions should be probabilistic functions of conjoint predicted utilities. Many probabilistic choice rules that can be applied at the individual level are now available (Green and Srinivasan 1978; Wind 1978; Green, Carroll and Goldberg 1981; Cattin and Wittink 1982; Finkbeiner 1988). They all recognize that each respondent will probably choose that brand having the highest predicted utility. Where the probabilistic choice rules differ is in how they estimate that probability. The estimate of share generated by each is the average estimated probability over the sample.

Despite the theoretical objections to the first choice rule, it has been found to work well in practice (Braun and Srinivasan 1975; Parker and Srinivasan 1976; Green and Srinivasan 1978; Huber and Moore 1979; Sudharshan, May and Shocker 1987; Elrod, Louviere and Krishna Kumar 1989). Interest in the topic was considerably increased by Finkbeiner's (1988) study comparing the first choice rule to five probabilistic rules. The first choice rule yielded the best estimates of share for both of two validation choice sets.

In this paper we explain why, despite the theoretical shortcomings of the first choice rule, it will in some circumstances predict share well. We investigate analytically the case in which conjoint analysis is used to predict share among two alternatives (which may be "buy the brand" versus "don't buy the brand"). For this case we are able to explain conceptually (and show analytically) that the first choice rule yields biased share predictions, as suspected by Louviere and Woodworth (1983). We also show that under some circumstances the bias will be considerable. Furthermore, we investigate one class of probabilistic rule and identify conditions under which it can yield nearly unbiased predictions of share. However, we also identify conditions under which the probabilistic rule can only yield more biased predictions of share than will the first choice rule.

We do not investigate analytically the case in which conjoint analysis is used to predict choice among more than two alternatives - due to its mathematical intractability - but we believe that our insights generalize to choice among an arbitrary number of alternatives. For example, we can explain Finkbeiner's (1988) findings, which we believe do not generalize to the prediction of share of choices made under substantially less controlled conditions. We conclude with some practical recommendations for users of conjoint simulators.

Predicting Share for Two Alternatives

We assume that all respondents choose that alternative with the highest utility, where utility for the i^{th} alternative is given by:

$$u_i = x_i b + e_i,$$

where x_i is a row vector describing the concept in terms of its attributes, b is a vector of part worths for the attributes, and e_i is an error term that captures departures from the conjoint model. We will refer to e_i as the choice error even though the "error" may lie in the model specification and not in the respondent's choice process.

For simplicity we will analyze the case in which choice is made from two alternatives. This case has the advantage of mathematical tractability, and we believe the insights generalize to choice among any number of brands. In the two-alternative case, we need estimate share for only the first alternative because the two shares must sum to one. Share for the first alternative is given by:

$$u = xb + e.$$

When both alternatives are brands, then u represents the difference in utilities for the brands $(u_1 - u_2)$, x the difference in the brand attribute vectors $(x_1 - x_2)$, and e the difference in errors $(e_1 - e_2)$. The respondent chooses the first brand whenever u is greater than zero.

However, the second alternative may be implicit, representing the alternative "don't choose the brand" or "choose some other brand." In this case, respondents are simply presented with one brand concept at a time and are asked whether or not they would buy it. This approach is useful when studying new concepts that have no clearly defined set of competing brands. It is possible to make "share" predictions in this case if the ratings or rankings used in estimation were also made against this implicit alternative.

True Share

Whether the second alternative is another brand or is implicit, true share for the first alternative is the fraction of respondents for which u is greater than zero. This fraction depends upon the distribution of u over the population of respondents, which is determined by the distributions of e and of b. We will assume that e is independently and identically distributed across respondents with mean zero and variance s_e^2. (The variance of the error term is likely to differ across respondents, but it can be assumed equal because u is defined only up to a ratio scale. When logit is used to estimate the part worths, the assumption of constant error variance is automatically satisfied.) Suppose further that the part worths (b) are distributed across respondents with mean b# and covariance matrix S_b. Then u will have mean xb# and variance:

$$s_u^2 = xS_bx' + s_e^2.$$

In words, this expression states:

variance in TRUE Utilities	=	heterogeneity variance	+	choice error variance

Finally, if the choice errors are normally distributed and the part worths multinormally distributed, then a closed-form expression for true share is easily derived. It is:

$$\text{TRUE SHARE} = \text{Phi}[xb\#/s_u],$$

where Phi[z] equals the probability that a standard normal random variable will be less than or equal to z.

First Choice Share

The first choice rule takes the fraction of respondents who have predicted utilities greater than zero as its estimate of share, where predicted utility is given by:

$$\begin{aligned} u^* &= xb^* \\ &= xb + x(b^* - b), \end{aligned}$$

where b* represents the estimated part worths. Note that the predicted utilities differ from the true utilities in the two respects discussed above: choice error is assumed not to exist, and the true part worths are replaced by estimates.

The fraction of respondents for whom u* is greater than zero depends upon the distributions of b and of (b*-b). If estimation of the part worths is at the individual level and is unbiased, then u* will have mean xb# and variance:

$$s_{u*}^2 = xS_bx' + xE[(b^* - b)(b^* - b)']x'$$

In words, this expression states:

variance in PREDICTED utilities	=	heterogeneity variance	+	estimation error variance

Furthermore, u* will be approximately normally distributed. This is so because xb is normally distributed and because x(b* - b) is distributed independently of xb and is approximately normally distributed. (The exact distribution of x(b* - b) depends upon the method of estimation.) The estimate of share obtained by the first choice rule is simply:

$$\text{FIRST CHOICE SHARE} \sim \text{Phi}[xb\#/s_{u*}],$$

where "~" means "is approximately equal to."

By comparing the expressions for true share and for first choice share we see that the first choice rule will yield a nearly unbiased estimate of true share if and only if the variance in true utilities (s_u^2) is nearly equal to the variance in predicted utilities (s_{u*}^2).

When the First Choice Rule Will Yield Nearly Unbiased Predictions of Share

Comparing the expressions for variance in true utilities and variance in predicted utilities allows us to say quite a bit about when the first choice rule will predict well.

<u>The first choice rule will yield nearly unbiased predictions of share whenever part worths are accurately estimated and when there is little error in the choice process</u>. When these conditions hold, both the variance in true utilities and the variance in predicted utilities will be very close to the heterogeneity variance, regardless of its size.

Choice error and estimation error will both tend to be small whenever:

*respondents are motivated to respond in a highly involved manner
*all attributes important to respondents are included in the profile description
*the concept description is readily comprehended by respondents
*the concept's combination of attributes is plausible

In addition, estimation error will tend to be small whenever:

*the estimates are based upon respondent reactions to many different profiles
*the profiles used in estimation vary in their attributes according to an orthogonal design
*only a few part worths parameters are estimated

Finally, there will be less error in the choice process (that is, the choice utilities will adhere closely to the conjoint model) whenever:

*the brands are accurately described by their profile descriptions
*the brands are similar to the profiles used in estimation

Many of these conditions are hard to meet in practice, and many are difficult to satisfy simultaneously. It is difficult to keep respondents highly motivated while asking them to respond to many profiles, and it is difficult to include all attributes important to respondents while estimating only a handful of part worths. Fortunately, there are two other conditions in which the first choice rule can predict well.

<u>The first choice rule will yield nearly unbiased predictions of share whenever respondents are very heterogeneous in their brand preferences</u>. Since the variance in true utilities and the variance in predicted utilities have heterogeneity variance in common, when heterogeneity variance is large any difference between the two utility variances will be small by comparison. While it is impossible to speak for all product categories, several authors have discovered appreciable heterogeneity in part worths (Pekelman and Sen 1974; Horsky and Rao 1984). Furthermore, the prevalence of substantial heterogeneity in consumer buying behavior toward existing brands is well established (Bass, Tigert and Lonsdale 1968; Bass 1974; Sabavala and Morrison 1977; Kalwani and Morrison 1977; Dalal, Lee and Sabavala 1984; Bass and Leone 1986; Elrod 1988). Finally, consumer heterogeneity serves as a theoretical explanation of product differentiation (Lancaster 1979).

While appreciable heterogeneity is likely to be found frequently in practice, respondents must be very heterogeneous to guarantee that first choice share will be nearly unbiased. It is hard to say at what point respondent heterogeneity guarantees negligible bias. However, it is always true that increased heterogeneity reduces bias in the first choice rule.

Finally, <u>the first choice rule can yield nearly unbiased predictions of share even when there is MUCH error in the choice process and part worths are INaccurately estimated</u>. This condition points out that the estimation and choice error variances can both be large provided that they are about equal in size. Then the predicted utilities will still have nearly the same variance as the true utilities. Note also that more accurate estimation in this case can lead to worse predictions of share using the first choice rule because it can make the two variances more unequal.

When the First Choice Rule Will Yield Substantially Biased Predictions of Share

There are two conditions under which the first choice rule will yield substantially biased predictions of share.

The first choice rule will yield biased share predictions whenever there is LITTLE error in the choice process and part worths are INaccurately estimated. When this is the case, then the variance in true utilities will be less than the variance in predicted utilities. As a consequence, in the case of two alternatives, first choice share will be between one half and the true share. We expect, in the case of more than two brands, that the first choice shares will have smaller variance across brands than will the true shares. More accurate prediction in this case would reduce the bias in the first choice share.

The first choice rule will yield biased share predictions whenever there is MUCH error in the choice process and part worths are ACCURATELY estimated. In this case, the variance in true utilities will be greater than the variance in predicted utilities. As a consequence, in the two-alternative case the true share will be between one half and the first choice share. And in the case of more than two brands, we expect the first choice shares will have larger variance across brands than will the true shares.

When this condition holds, more accurate estimation of the part worths can only increase the bias in the first choice share. However, as we now show, probabilistic prediction can lessen bias in this condition.

When Probabilistic Prediction at the Individual Level Will Improve Share Prediction Relative to the First Choice Rule

One alternative to the first choice rule is to make probabilistic predictions of choice at the individual level and then take as the estimate of share the average probability of choice over the sample of respondents. We consider one type of probabilistic rule applied at the individual level--one which is equivalent to adding random error to the predicted utilities. Applied at the individual level (for given b*), the probability of choosing the brand is given by:

$$p = \text{Prob}[u^{\bullet} > 0 \mid b^*],$$

where $u^{\bullet} = xb^* + o$, p is the predicted probability of choice at the individual level and o is the added random error with mean zero and variance s_o^2. For lack of better terms, we will refer to o as the simulated error and to $u^{\bullet}$ as the simulated utility. If the simulated error is assigned zero variance, then the simulated utilities are equal to the predicted utilities and the probabilistic rule is equivalent to the first choice rule (with $p = 1$ if $xb^* > 0$ and $p = 0$ if $xb^* < 0$). As the variance in the simulated error is increased, the value of p moves away from the 0-1 extremes. Where probabilistic rules of this type can differ among themselves is in the exact choice of distribution for o, but a key decision is the size of its variance--the smaller the variance, the closer the probabilistic rule is to the first choice rule. This effect is clearly demonstrated by Johnson (1988), where multiplying ACA Model 2 predicted utilities by scaling factors greater than one (holding simulated variance constant) is equivalent in our treatment to decreasing the simulated variance while holding the predicted utilities constant. ACA Model 2 assigns the logistic distribution to the simulated errors. Another alternative is to have the simulated errors be normally distributed. The choice between these two error specifications should make little difference in practice.

The probabilistic share is the average value of p over the sample of respondents, which is the same as $\mathrm{Prob}[xb^* + o > 0]$, where b^* is now recognized as being a random variable. If the simulated error variance is designated s_o^2, then the simulated utilities across all respondents will have mean $xb\#$ and variance:

$$s_{u^*}^2 = xS_bx' + xE[(b^* - b)(b^* - b)']x' + s_o^2.$$

In words, this expression states:

variance in SIMULATED utilities	=	heterogeneity variance	+	estimation error variance	+	simulated error variance
	=	variance in PREDICTED utilities			+	simulated error variance

If the simulated errors are normally distributed, then the simulated utilities are nearly normally distributed and probabilistic share is given by:

$$\text{PROBABILISTIC SHARE} \sim \mathrm{Phi}[xb\#/s_{u^*}].$$

Now we are in a position to say when this type of probabilistic rule can be expected to yield better share estimates than the first choice rule. We first consider the circumstances in which this type of probabilistic rule will be more biased.

Probabilistic rules of this type will yield MORE biased estimates of share than the first choice rule whenever the variance in predicted utilities already exceeds the variance in true utilities. This will be the case whenever there is LITTLE error in the choice process and part worths are INaccurately estimated. The reason for this conclusion is that there is already too much error in the predicted utilities, so adding simulated error to them can only increase bias.

These conditions are likely to hold for many of the choice tasks used to assess the validity of conjoint models. As noted by Louviere (1988), most validity studies of conjoint analysis have taken as their criterion the ability to predict respondent response to holdout profiles presented at the end of the conjoint estimation session. Examples are Green, Rao and DeSarbo (1978); Huber and Moore (1979); Jain, Acito, Malhotra and Mahajan (1979); Finkbeiner (1988); and Elrod, Louviere and Krishna Kumar (1989). Johnson's (1988) comment on Finkbeiner's study is applicable to most validity studies of conjoint analysis: "I am not surprised that when respondents are shown concepts described on a few attributes, most of them are able to select the concepts having highest utility for them. Indeed, if one concept dominates the others so clearly as in this 4-concept example, the choice task may be little more than a reading comprehension test" (p. 108). Thus choice error in these studies can be expected to be quite small. Whether it is smaller than the estimation error for the typical conjoint study is harder to determine. The more ambitious the study in terms of the number of attributes and attribute levels, the more likely it is that the estimation error is larger than the choice error and that the first choice rule will understate differences in share across brands.

There are conditions in which judicious choice of the simulation error variance can yield a nearly unbiased estimate of share, even when the first choice share is substantially biased.

Probabilistic rules can yield LESS biased estimates of share than the first choice rule whenever the variance in predicted utilities is less than the variance in true utilities. This will be the case whenever there is MUCH error in the choice process and part worths are ACCURATELY estimated. In this case, there is too little variance in the predicted utilities, and adding simulated error of the correct variance can equate the variances in simulated and true utilities and yield nearly unbiased estimates of share. However, reduction in bias is not guaranteed because it is possible to add too much simulation variance.

This condition of having too little variance in predicted utilities relative to the true choice utilities is likely to hold when predicting share for choice tasks different from the conjoint task used to estimate part worths. Asking respondents to react to a brand rather than a profile at a different time in a different context would all cause the utilities governing choice to deviate from the conjoint model. This is particularly likely to be true when predicting choice in the marketplace. Johnson provides insight on this matter when he notes that, in his experience, probabilistic rules predict market shares better than the first choice rule. He speculates as to why this might be so: "Everyone would agree that a buyer's actual purchase decision is affected by factors other than his preference at some previous point in time. It is also affected by out-of-stock conditions, point-of-sale promotions, effects of advertising, appetite for variety, and misperceptions of product attributes, to name a few factors" (p. 107). Recognition of the importance of these outside factors is widespread (Pessemier, Burger, Teach and Tigert 1971; Morrison 1979; Finkbeiner 1988). All of these factors would cause the utilities governing choice to depart substantially from the conjoint model. In our terms, choice error variance will be large and the first choice rule will tend to overstate differences in share across brands.

Some Practical Guidelines for Users of Choice Simulators

The insights described in this paper may well be of interest to designers of conjoint choice simulators, but it is always fair to ask what the implications are for users of these simulators. We can think of several implications.

Often the choice of simulator will make little difference in the conclusions reached by managers. As noted by Johnson (1988), Finkbeiner's comparison of 5 simulators showed that all of them correctly predicted the rank order of the true shares for the 4-concept choice task. If the purpose of the conjoint study is to identify that concept that will have the greatest appeal in the marketplace, then the different choice simulators are likely to agree on what that concept is. Probably the best practice is to rely on the conjoint analysis to identify several appealing concepts to be considered further by management. Different conjoint choice simulators are likely to yield nearly the same set of concepts.

The profiles used in simulation should be designed so that the profile with the largest share is the one with the most appeal to management. A number of writers have warned that an unrestricted search for the concept having the highest share will not identify the most desirable concept from management's point of view unless costs are taken into account (Bachem and Simon 1981; Hauser and Simmie 1981; Sen 1982; Gavish, Horsky and Srikanth 1983). For example, a $5000 luxury car might have great appeal to respondents, but management has no interest in making something that must be sold at a loss.

There are at least two ways to take costs into account when identifying desirable concepts. One way is after the prediction of shares. The researcher would have to predict share for all concepts and then, taking costs into account, predict the dollar profit to be made per unit for each concept. The most desirable concepts are those with the largest values for share times unit profit. This approach has two drawbacks. First, the choice simulator cannot be used to eliminate concepts based on share alone. Second, the estimates of share must be ratio scaled. We have seen that conjoint choice simulators, while they tend to agree on which concepts have the highest share, differ in their predictions about the relative sizes of these shares across concepts. The probabilistic rule will predict smaller differences in shares across concepts than will the first choice rule.

We recommend a second method, which is to begin by specifying all concepts to be evaluated by the simulator in terms of their non-price attributes. Then predict the cost for each concept and set the price level for each concept at some specified dollar amount above cost. This is done most easily by estimating the cost of providing each level on each attribute (Green, Helsen and Shandler 1988). Then the concepts with highest predicted share will be the most desirable from management's point of view because the same dollar amount will be made per unit for each concept. And the choice simulators do not differ very much in their predictions about which concepts have the highest share.

One potential difficulty with implementing our recommendation is that the price levels that equate unit profits will not equal the tested price levels, and some cases may not fall within the range of prices tested. A combination of interpolation and (more tentatively) extrapolation must be used to predict respondent utilities for concepts at the profit-equalizing prices.

<u>Know what you want your share predictions to mean</u>. After discussing the many factors that can cause market shares to differ from conjoint analysis predictions, Johnson goes on to note that "not everyone agrees about the extent to which allowances should be made for these factors in conjoint simulators. One school of thought holds that the conjoint simulator should be concerned with preference alone, and that the most appropriate test is with holdout concepts" presented at the end of the conjoint estimation task (p. 107). This school of thought argues that conjoint analysis is concerned only with the product (and usually price) elements of the marketing mix, and no attempt should be made to include other determinants of choice behavior. To use a consumer product as an example, choice shares for holdout concepts presented at the tail end of the conjoint estimation task give the researcher some idea of how consumers will behave when presented with these two brands in the same store at the given prices and with everything else held constant. However, the choice shares will tend to overstate the differences in market shares for these brands, even when the brands do not differ <u>at the aggregate level</u> on the other factors that influence choice in the marketplace. For example, two brands can have equal distribution in aggregate and yet each will appear in some stores without the other. Equal distribution for two brands in aggregate does not mean that every consumer will always be presented with both brands side by side. Since availability influences choice, the result will be that the market shares for the two brands will be more nearly equal than the choice shares. Similarly, two brand managers can spend the same amount on dealing in aggregate and

have their brands appear on sale an equal percentage of the time. Yet, since few consumers will be presented with both brands on deal at the same time in the same stores, dealing will cause the market shares to be more nearly equal than the choice shares.

Understood this way, we believe many managers will want a prediction of market share based on product attributes and price while other factors are held constant <u>in aggregate</u> but not necessarily for every potential buyer in the population. Some of these other factors are under management control, which is exercised at the aggregate (or perhaps segment) level, and not usually at the level of the individual customer. Other factors are not under management control but can be forecasted and taken into account, again usually at the aggregate level. Managers typically have little knowledge or control about the myriad of alternatives and influences confronting individual customers, and predictions of share based on the assumption that these factors are identically constant for all potential customers will usually have limited meaning and usefulness. Recognizing the importance of these other factors on brand choice can avoid overstating the differences in share to be expected by differences in product or price alone.

Wanting share predictions that take into account the expected impact of non-product/ non-price influences is not the same as getting such predictions (Cattin and Wittink 1982). Averaging probabilistic predictions made at the individual level is one way to capture the role these other factors are likely to play while holding them constant in aggregate across brands. Probabilistic predictions can be based upon supplementary judgments by respondents about probability of purchase as a means for scaling utilities, as in ACA (Adaptive Conjoint Analysis System, Johnson 1987). Alternatively, the conjoint estimates can be based on ratings expressed in terms of likelihood of purchase (as, for example, in Elrod, Louviere and Krishna Kumar 1989). Either way, respondents provide information about the likelihood that other influences not tapped in the conjoint task will cause a change in their choice at the time of purchase. Some writers have recognized that the first choice rule will work well whenever influences other than product and price are relatively unimportant (Shocker and Srinivasan 1979; Sudharshan, May and Shocker 1987), as is more often true with high-involvement brand choice decisions. We have shown analytically that this belief is well founded. As other influences have a greater impact, the first choice shares will be too extreme in their predictions relative to the true shares, and probabilistic predictions are likely to be less biased. Once we appreciate this, we are no longer surprised that the first choice rule does well in predicting share among choices made in experimental settings designed to minimize the influence of other factors. We have good reason to believe that share of choices made in less controlled settings can be better predicted by probabilistic rules, even though the effect of other influences on brand choice has been held constant in aggregate across brands.

Acknowledgments

Our thinking on this topic benefited greatly from conversations with, and comments from, Carl T. Finkbeiner, Richard M. Johnson, Wagner A. Kamakura and Jordan J. Louviere. We accept responsibility for any errors.

References

Albers, Sonke (1979), "An Extended Algorithm for Optimal Product Positioning," <u>European Journal of Operations Research</u>, 3, 222-231.

Bachem, Achem and Herman Simon (1981), "A Product Positioning Model with Costs and Prices," European Journal of Operations Research, 7 (August), 362-370.

Bass, Frank M. (1974), "The Theory of Stochastic Preference and Brand Switching," Journal of Marketing Research, 11 (February), 1-20.

Bass, Frank M. and Robert P. Leone (1986), "Estimating Micro Relationships from Macro Data: A Comparative Study of Two Approximations of the Brand Loyal Model Under Temporal Aggregation," Journal of Marketing Research, 23 (August), 291-297.

Bass, Frank M., Edgar A. Pessemier, and Donald R. Lehmann (1972), "An Experimental Study of the Relationship Between Attitudes, Brand Preference, and Choice," Behavioral Science, 17 (November), 532-541.

Bass, Frank M., Douglas J. Tigert, and Ronald T. Lonsdale (1968), "Market Segmentation: Group Versus Individual Behavior," Journal of Marketing Research, 5 (August), 264-270.

Braun, Michael A. and V. Srinivasan (1975), "Amount of Information as a Determinant of Consumer Behavior Toward New Products," 1975 Combined Proceedings, Chicago: American Marketing Ass., 373-378.

Cattin, Philippe and Dick R. Wittink (1982), "Commercial Use of Conjoint Analysis: A Survey," Journal of Marketing, 46 (Summer), 44-53.

Dalal, S. R., J. C. Lee and J. Sabavala (1984), "Prediction of Individual Buying Behavior: A Poisson-Bernoulli Model with Arbitrary Heterogeneity," Marketing Science, 3 (Fall), 352-367.

Elrod, Terry (1988), "A Management Science Assessment of a Behavioral Measure of Brand Loyalty," Advances in Consumer Research, 15, 481-486.

Elrod, Terry, Jordan J. Louviere and S. Krishna Kumar (1989), "How Well Can Conjoint Analysis Predict Choices from Efficient Choice Sets?" working paper, Vanderbilt Univ., May.

Finkbeiner, Carl T. (1988), "Comparison of Conjoint Choice Simulators," in Proceedings of the Sawtooth Software Conference on Perceptual Mapping, Conjoint Analysis, and Computer Interviewing, Sun Valley ID: Sawtooth Software, 75-103.

Gavish, Bezalel, Dan Horsky and Kizhanatham Srikanth (1983), "An Approach to the Optimal Positioning of a New Product," Management Science, 29 (November), 1277-1297.

Green, Paul E., J. Douglas Carroll and Stephen M. Goldberg (1981), "General Approach to Product Design Optimization Via Conjoint Analysis," Journal of Marketing, 45 (Summer), 17-37.

Green, Paul E., Wayne S. DeSarbo and Pradeep K. Kedia (1980), "On the Insensitivity of Brand-Choice Simulations to Attribute Importance Weights," Decision Sciences, 11 (1980), 439-450.

Green, Paul E., Kristiaan Helsen and Bruce Shandler (1988), "Conjoint Internal Validity Under Alternative Profile Presentations," Journal of Consumer Research, 15 (December), 392-7.

Green, Paul E. and Pradeep K. Kedia (1983), "A Method for Weighting Respondents in Conjoint-Based Computer Choice Simulators," Wharton School, Working Paper No.83-006R, August.

Green, Paul E., Vithala R. Rao and Wayne S. DeSarbo (1978), "Incorporating Group-Level Similarity Judgments in Conjoint Analysis," Journal of Consumer Research, 5 (December), 187-193.

Green, Paul E. and V. Srinivasan (1978), "Conjoint Analysis in Consumer Research: Issues and Outlook," Journal of Consumer Research, 5 (September), 103-123.

Hauser, John R. and Patricia Simmie (1981), "Profit Maximizing Perceptual Positions: An Integrated Theory for the Selection of Product Features and Price," Management Science, 27 (January), 33-56.

Horsky, Dan and M. R. Rao (1984), "Estimation of Attribute Weights from Preference Comparisons," Management Science, 30 (July), 801-822.

Huber, Joel and William Moore (1979), "A Comparison of Alternative Ways To Aggregate Individual Conjoint Analyses," in 1979 Educators' Conference Proceedings, Chicago: American Marketing Ass., 64-68.

Jain, Arun K., Franklin Acito, Naresh K. Malhotra and Vijay Mahajan (1979), "Comparison of the Internal Validity of Alternative Parameter Estimation Methods in Decompositional Multiattribute Preference Models," Journal of Marketing Research, 16 (August), 313-322.

Johnson, Richard M. (1987), "Adaptive Conjoint Analysis," unpublished manuscript, Sun Valley ID: Sawtooth Software.

Johnson, Richard M. (1988), "Comment [on 'Comparison of Conjoint Choice Simulators,' by Carl Finkbeiner]," in Proceedings of the Sawtooth Software Conference on Perceptual Mapping, Conjoint Analysis, and Computer Interviewing, Sun Valley, ID: Sawtooth Software, 105-108.

Kalwani, Manohar U. and Donald G. Morrison (1977), "Parsimonious Description of the Hendry System," Management Science, 23 (January), 467-477.

Lancaster, Kelvin (1979), Variety, Equity and Efficiency, New York: Columbia Univ. Press.

Louviere, Jordan J. (1988), "Conjoint Analysis Modelling of Stated Preferences: A Review of Theory, Methods, Recent Developments and External Validity," Journal of Transport Economics and Policy, (January), 93-119.

Louviere, Jordan J. and George Woodworth (1983), "Design and Analysis of Simulated Consumer Choice or Allocation Experiments: An Approach Based on Aggregate Data," Journal of Marketing Research, 20 (November), 350-367.

McAlister, Leigh and Edgar Pessemier (1982), "Variety Seeking Behavior: An Interdisciplinary Review," Journal of Consumer Research, 9 (December), 311-322.

Moore, William L. (1980), "Levels of Aggregation in Conjoint Analysis: An Empirical Comparison," Journal of Marketing Research, 17 (November), 516-523.

Morrison, Donald G. (1979), "Purchase Intentions and Purchase Behavior," Journal of Marketing, 43 (Spring), 65-74.

Parker, Barnett R. and V. Srinivasan (1976), "Consumer Preference Approach to the Planning of Rural Primary Health-Care Facilities," Operations Research, 24 (September-October), 991-1025.

Pekelman, Dov and Subrata K. Sen (1974), "Mathematical Programming Models for the Determination of Attribute Weights," Management Science, 20 (April), 1217-1229.

Pessemier, Edgar, Philip Burger, Richard Teach and Douglas Tigert (1971), "Using Laboratory Brand Preference Scales to Predict Consumer Brand Purchases," Management Science, 17 (February), B371-B385.

Reibstein, David J. (1978), "Prediction of Individual Probabilities of Brand Choice," Journal of Consumer Research, 5 (December), 163-168.

Sabavala, Darius J. and Donald G. Morrison (1977), "A Model of TV Show Loyalty," Journal of Advertising Research, 17 (December), 35-43.

Sen, Subrata K. (1982), "Issues in Optimal Product Design," in Analytic Approaches to Product and Marketing Planning, edited by Rajendra K. Srivastava and Allan D. Shocker, Cambridge MA: Marketing Science Institute, 265-273.

Shocker, Allan D. and V. Srinivasan (1979), "Multiattribute Approaches for Product Concept Evaluation and Generation: A Critical Review," Journal of Marketing Research, 16 (May), 159-180.

Silk, Alvin J. and Glen L. Urban (1978), "Pre-Test-Market Evaluation of New Packaged Goods: A Model and Measurement Methodology," Journal of Marketing Research, 15 (May), 171-191.

Sudharshan, D., Jerrold H. May and Allan D. Shocker (1987), "A Simulation Comparison of Methods for New Product Location," Marketing Science, 6 (Spring), 182-201.

Urban, Glen L. and John R. Hauser (1980), Design and Marketing of New Products, Englewood Cliffs NJ: Prentice-Hall.

Wiley, James B. and James T. Low (1983), "Monte Carlo Simulation Study of Two Approaches for Aggregating Conjoint Data," Journal of Marketing Research, 20 (November), 405-417.

Wind, Yoram (1978), "Issues and Advances in Segmentation Research," Journal of Marketing Research, 15 (August), 317-337.

Wittink, Dick R. and David B. Montgomery (1979), "Predictive Validity of Trade-off Analysis for Alternative Segmentation Schemes," in 1979 Educators' Conference Proceedings, Chicago: American Marketing Ass., 69-73.

Assessing the Validity of Conjoint Analysis

Richard M. Johnson
Sawtooth Software, Inc.

Researchers often use "hold-out concepts" to assess the validity of conjoint analysis. The respondent ranks or rates concept statements for overall preference. Those responses are not used in the estimation of his part worths. They are assumed to reveal his true preferences, and the validity of conjoint analysis is assessed by how well it predicts those preferences.

However, some researchers have argued that if conjoint analysis fails to predict responses to hold-out concepts, that failure may be due as much to errors in response to the concepts as to errors in the conjoint predictions.

In many conjoint studies a far greater investment is made in the quality of the conjoint estimates than in the hold-out concept responses. The respondent may spend 20 minutes providing data for conjoint estimation, but only a few seconds providing data on hold-out concepts.

There are at least two sources of potential error in responses to hold-out concepts:

1. Position Effects: respondents might pay greater attention to attributes with more prominent positions in the concept statements, such as those near the top rather than in the middle or bottom of the concept statements.

2. Unreliability: respondents might give different answers if asked the same questions again.

This paper presents data showing that much of the apparent error in conjoint predictions can be attributed to these types of errors in responses to hold-out concepts.

Previous Evidence:

Johnson (Johnson, R. M.. "Problems in Applying Conjoint Analysis," Conference on Analytic Approaches to Product and Marketing Planning, Vanderbilt University, October 1981) reported results of a full-profile conjoint study of 7 attributes, each with 3 levels, where the concern was whether respondents would pay as much attention to product attributes described toward the bottom of the concept statements as to those near the top.

An orthogonal design was used to construct 18 concept cards. Two versions of the questionnaire were used, identical except for order of presentation of the attributes on the cards. We refer to the attributes with letters A through G. In Version 1 the attributes appeared in order A through G. Version 2 presented them in reverse order. Attribute G was represented by pictures, and all other attributes by phrases.

Several hundred respondents were given shuffled decks of 18 cards and asked to sort them into 7 categories ranging from "most likely to buy" to "least likely to buy." Categories were coded with values of 7 through 1. The average score was computed for each card in each version, and ordinary least squares (OLS) regression was used to obtain average part worth values for each version. The range in part worths for each attribute was taken as a measure of its relative importance, after normalizing to sum to 100.

Average Attribute Importances

Attribute	Version 1 (ABCDEFG)	Version 2 (GFEDCBA)	Difference
A	33.5	26.1	7.4
B	24.4	22.9	1.5
C	15.5	17.2	-1.7
D	14.3	15.0	-0.7
E	4.2	5.2	-1.0
F	2.3	7.6	-5.3
G	5.8	6.0	- .2

Attribute G, represented by a picture rather than words, was found to be about equally important whether it appeared at the top or the bottom of concept statements.

When Attribute A was presented at the top of a concept its relative importance was 33.5; when presented at the bottom its relative importance was 26.1. The difference of 7.4 is apparently a position effect. Attribute F had an importance of 7.6 when it was the at the top of those attributes expressed in words, and 2.3 when it was at the bottom of those expressed in words, for a position effect of 5.3.

Attributes B and E, which were either next to the top or next to the bottom of the attributes expressed in words, showed effects 1.0 or 1.5.

Attributes C and D, which were never near the top or the bottom, were each found to be marginally more important when in the fourth position among the attributes expressed in words rather than third position.

These data indicate that reactions to concept statements, and estimates of part worths for the attributes, can depend strongly on the relative positions of attributes in the concepts.

The Experiment:

To further quantify these factors, a methodological component was added to a survey conducted in the Spring of 1989 among users of Sawtooth Software's Ci2 System for computer interviewing. The survey was on diskettes mailed to slightly over 300 Ci2 users. It contained two portions of interest:

> There was a conjoint interview (using Sawtooth Software's ACA System for Adaptive Conjoint Analysis) to estimate respondents' part worths for attributes.
>
> Following the conjoint interview, 8 hold-out concepts were presented, described by the same attributes. Respondents' level of interest in each was asked on a 100 point scale.

We requested responses from individuals who had experience with Ci2 and who also had influence in the decision to acquire interviewing software. No incentive was offered, other than that of influencing the next version of Ci2, and no reminder notices were sent. After 6 weeks, 45% of the survey disks had been returned, with an average of 1.25 complete interviews per disk. The results reported here are based on the first 175 complete interviews received.

Position Effects:

The hold-out concepts were presented in two versions, with random selection of version for each respondent. Approximately half the respondents received concepts with attributes presented in order A through K. The remainder of the respondents received attributes in order GHIJ ABCDE K. Attribute K, price, was the last attribute in both versions. All attributes were represented in words.

Respondents' part worths were known for the attribute levels in each concept. We might expect that a respondent's ratings for the concepts could be predicted by adding the appropriate part worths. On the other hand, if the importance of an attribute depends on relative position in the concept statement, then attributes in prominent positions would be expected to receive more weight in the formation of preferences.

An OLS regression analysis was done for each version. Each analysis had about 700 observations, 700 being approximately the number of respondents receiving that version times the number of concepts rated by each respondent.

The 11 independent variables were part worths estimated by ACA for each attribute as it appeared in each concept. The dependent variable consisted of ratings of each concept by each respondent, after normalizing each respondent's ratings to sum to 100.

If respondents' preferences were determined solely by the part worths estimated by ACA then all attributes would be expected to receive nearly equal regression weights. On the other hand, if attributes in prominent positions received more weight when respondents rated concepts, then one might expect the regression weights of attributes in prominent positions to exceed those for attributes in less prominent positions. The results were as follows:

Regression Weights For Attributes
(shown in same order as presentation in concepts)

Version 1			Version 2		
Attribute	Weight	t value	Attribute	Weight	t value
A	5.3	3.73	G	-0.1	-0.02
B	3.9	1.87	H	2.4	1.21
C	15.1	7.80	I	0.5	.26
D	5.3	2.21	J	-8.6	-3.74
E	-5.5	-2.19			
F	2.5	1.00	A	7.3	4.90
			B	0.4	.16
G	1.8	.67	C	12.7	7.58
H	0.2	.10	D	-0.2	-0.10
I	-5.8	-2.95	E	7.2	3.20
J	-1.4	-.65	F	-2.0	-0.71
K	5.5	2.71	K	7.0	3.61

The price attribute (K), which had the same position in both versions, had similar importances in the two versions.

Attributes D, E, I, and J had large and significant differences between versions. However, there seems to be no pattern to these differences. Responses to these concepts apparently did depend strongly on relative positions of attributes in the concept statements, although in a less predictable way than in the earlier example.

Attribute C received a large and highly significant weight in both versions, and attribute J received a negative weight in both versions. We should probably not make too much of this, in view of the large position effects, but responses to the concepts may be capturing a different kind of information than responses to the pairwise tasks in the ACA interview. (In fact, regression weights such as these can be used after the fact to correct for redundancy in attributes included in a pairwise conjoint study. Weights of similar attributes are reduced, and two identical attributes would receive weights of half what either would receive by itself.)

There are strong position effects, but they are chaotic rather than regular. This may be due to the greater complexity of these concepts, with 11 rather than just 7 attributes.

The question arises of how serious the distortion created by such position effects may be. To quantify this, correlations were computed between preference and <u>unit-weighted</u> part worths for respondents receiving each version. The complements of squared correlations indicate relative proportions of error variance. These may be contrasted to similar values for the regressions yielding the weights described above:

Relative Error Variances
(1 - R Squared)

	Unit Wts	OLS Wts	Ratio
Version 1	.897	.763	.851
Version 2	.919	.759	.826

The last column of this table shows the reduction in relative error variance obtained by incorporating position effects in the prediction of concept ratings. Results for the two versions are quite similar, indicating that position effects in responses to hold-out concepts account for somewhere in the neighborhood of 16 percent of the total error variance of the conjoint predictions.

These findings have strong implications for research based on ratings of complex concepts. When concepts must be rated, it seems desirable to have alternate versions so that position effects can be identified. In fact, it would be desirable to have <u>many</u> versions so that position effects could be unconfounded from other effects. Even then, if position effects are large, it is not clear how to avoid errors at the individual level.

Reliability and Predictability of Hold-out Concept Ratings:

In the hold-out concepts, Concept 1 presented a product much like the current version of Ci2 at a price of $3,000. Concept 2 presented an improved product with greater capability on every attribute and a price of $5,000. By any standard, Concept 1 and Concept 2 were "quite different" from one another.

Concept 3 through Concept 6 were extremely varied, ranging from "simpler" than Concept 1 to "fancier" than Concept 2. The main purpose for Concepts 4 through 6 was to obscure respondents' memories of how they had rated the first two.

Concept 7 was identical to the Concept 2, and Concept 8 was identical to Concept 1.

Since Concepts 1 and 2 were rated at nearly the same time by each respondent, we can regard those ratings as a paired comparison. Similarly, we can regard the rating of Concepts 8 and 7 as a replication of the paired comparison.

The simplest approach to measuring reliability is through an examination of 2 x 2 tables. Respondents with tied ratings in either comparison were deleted for this analysis. Of the 141 remaining, here are the percentages choosing each alternative in each replication:

Reliability of Hold-out Concept Preference

		Replication 2: Preferred Concept 1	Replication 2: Preferred Concept 2	Total
Replication 1	Preferred Concept 1	52%	10%	62%
	Preferred Concept 2	10%	28%	38%
	Total	62%	38%	100%

Concept 1 was quite strongly preferred, with 62% of respondents preferring it in either replication. However, 20% of the respondents revealed inconsistent preferences.

Since concepts were specified in terms of the attributes studied in the conjoint section of the interview, it was possible to model each respondent's preference based on his part worths. Here are results on the accuracy of conjoint predictions using the "first choice rule."

Accuracy of Prediction of Replication 1

	Preferred Concept 1	Preferred Concept 2	Total
Predicted Concept 1	47%	13%	60%
Predicted Concept 2	15%	25%	40%
Total	62%	38%	100%

Accuracy of Prediction of Replication 2

	Preferred Concept 1	Preferred Concept 2	Total
Predicted Concept 1	48%	12%	60%
Predicted Concept 2	14%	26%	40%
Total	62%	38%	100%

The conjoint predictions are a little less accurate than those made by predicting either replication from the other, but not dramatically less so. We would be wrong an average of 27% of the time when predicting concept choice from the part worths, but we would be wrong 20% of the time predicting either replication from the other. Using this simple way of accounting for error, a large proportion of the error of prediction appears to be due to unreliability in the concept preferences.

Another approach to measurement of reliability is through correlation analysis. For each replication we computed the difference between ratings of concepts 1 and 2, and the difference between ratings of concepts 8 and 7. Ties result in differences of zero, and comparisons involve all 175 respondents. The "test-retest reliability" correlation between these two preference measures was .667.

As a measure of validity of conjoint predictions, we can correlate the difference in predicted share of preference for concepts (using a logit model) with the difference in ratings of concepts. The correlations are .527 and .536 for replications 1 and 2 respectively, with an average value of .532.

Psychometricians use a "correction for attenuation" to estimate the correlation that would be observed between two variables if either or both were measured with perfect reliability. This is just the observed correlation between the two variables divided by the product of the square roots of their reliability correlations. We lack a measure of reliability for the conjoint predictions but we do have a reliability measure for concept choice. The estimate of the correlation between conjoint predictions and a perfectly reliable measure of conjoint choice would be:

$$.532 / \text{sqrt}(.667) = .651.$$

The relative error variance of predictions expected if concept preference were completely reliable is:

$$(1 - .651 \text{ squared}) = .576.$$

The relative error variance actually observed in the conjoint predictions is:

$$(1 - .532 \text{ squared}) = .717.$$

The proportion of the error variance due to unreliability in concept preference is therefore

$$(.717 - .576) / .717 = .197$$

or about 20%.

Summary

We have seen that when the validity of conjoint analysis is assessed by its ability to predict hold-out concepts, much of the error variance is actually due to problems with the concept data rather than problems with the conjoint data.

1. There are strong position effects in respondents' reactions to hold-out concepts. In a simple case with few attributes, those near the top of concept statements were found to be more important. In more complex concepts involving 11 attributes, position effects were also strong but no clear pattern emerged.

2. It is possible to infer the proportion of the total error variance of conjoint predictions which is in fact due to position effects in respondents' reactions to the concept statements. In this example approximately 16% of the error variance of the conjoint predictions could be attributed to this source.

3. When respondents are asked to repeat choice tasks with the same concepts, they do not always give the same answer a second time. In this example 20% of the respondents gave contradictory responses when choosing between two concepts that were "quite different" from one another. For comparison, predicting concept choice from conjoint data using the "first choice rule" yielded an only slightly higher error rate of only 27%.

4. It is possible to infer the proportion of total error variance of conjoint predictions which is in fact due to unreliability of respondents' ratings of concepts. In this example approximately 20% of the error variance of the conjoint predictions could be attributed to this source.

5. When hold-out concepts are used to assess the validity of conjoint analysis, it is important to:

(a) have alternate versions of the concepts (presenting attributes in different orders) to assess the magnitude of position effects and to avoid confounding position effects with substantive issues.

(b) have within-respondent replication of some concepts to assess the extent to which apparent errors of prediction may in fact be due to unreliability of concept ratings.

(c) undertake an analysis to determine the extent to which apparent errors of prediction are in fact due to each of these sources.

NEW FINDINGS WITH OLD DATA USING CLUSTER ANALYSIS

Tara Thomas
Blue Cross and Blue Shield of Iowa

INTRODUCTION

For the past four years, Blue Cross and Blue Shield of Iowa (BCBSI) has been engaged in extensive consumer research. The purposes have been to uncover how consumers make choices in a complicated decision environment and how they evaluate the component features of insurance programs. Of particular interest has been the influence of service oriented activities on the choices that consumers make. The eventual goal is to determine to what extent choices can be altered by the inclusion or exclusion of service features in health insurance product designs.

During the summer of 1987, BCBSI and IMR Systems, Ltd. began a process that ended with a useful and informative perceptual mapping study of the image of BCBSI in the Iowa marketplace. The research would have been useful to us if the study had ended with that final report. But the true value of the research has been realized with continued analysis of this and other data sets, using two separate clustering methodologies.

The purpose of clustering these data sets was to investigate the possibility that market segments exist for health insurance products. As we recognized the possibility of segments within the data sets under investigation, we began to discuss the strategic implications of these results. Identification and description of the market segments would facilitate product design and distribution, and allow us to position and promote these products appropriately. In addition, understanding consumer choice patterns might eventually allow BCBSI to "custom design" products in a way that would enhance the composition of our insurance risk pools.

Research Approach

During the summer and fall of 1988, BCBSI undertook a critical reexamination of five different data sets. The purposes of this analysis were to determine whether or not definable market segments exist and whether segment preferences are reflected in product choices. Four of these data sets were created through ACA (ACA System for Adaptive Conjoint Analysis), the fifth through APM (APM System for Adaptive Perceptual Mapping). Each represented a distinct set of respondents facing very different health insurance decisions. One ACA study examined only service features of health insurance, while the others included both service and insurance attributes.

To uncover possible market segments, the data were clustered using two different procedures: an attribute cluster approach using SPSS/PC+ and a respondent cluster approach using CCA (CCA System for Convergent Cluster Analysis). Despite the fact that none of the original studies were designed as segmentation studies, the findings were comparable across all five. Across the data sets, two distinct segments emerge in the health insurance industry: one that prefers traditional health insurance products, and one that prefers the features of HMO health insurance products.

Findings and Strategic Implications

These two segments -- the "Traditional" and "HMO" segments -- are comparable across all five data sets. Quite frankly, this is not the result that we expected. In addition, several data sets exhibit what we believe to be initial indications of other emerging segments in the market. Our interpretation is that the market is in transition; new segments will emerge as consumers become exposed to wider choices in health insurance and gain better understanding of their decision environment.

Once the segments were identified using cluster analysis, the data sets were separated to conduct simulation analyses. The purpose of running the simulations was to determine which BCBSI product might appeal to different market segments. Not surprisingly, and quite reassuringly, those consumers who preferred traditional insurance were attracted to traditional products. And consumers attracted to HMO (Health Maintenance Organization) style insurance preferred HMO products. This indicates to us that consideration of product development for a segmented market was in order for BCBSI.

As a final step, the service features of products were excluded from a second set of simulation analyses. When these service features were excluded, consumer preferences changed -- sometimes dramatically. A substantial shift in preference away from the HMO alternatives (which are grounded in high service and convenience) toward the traditional insurance products was apparent for all segments. This indicated that consumers often select their health insurance coverage (when they have a choice) based on how important service is to them; many consumers were willing to give up their choice of physician and other freedoms associated with traditional insurance to obtain enhanced service.

STUDY BACKGROUND AND DATA BASES

Background Literature

Our interest in conducting this health insurance segmentation study was initially raised by an article by Woodside, Nielsen, Walters, and Muller published in the Journal of Health Care Marketing. The authors identified four distinct and identifiable segments in the market for hospital services: the "Old-Fashioneds," "Value Conscious," "Affluents," and "Professional Want-It-Alls." These four segments seek different benefits from the hospitals they prefer, and have characteristic demographic profiles. The study confirmed the contention of Kotler and Clarke that consumers seek different benefits from health care. Some seek "Quality," some "Service," some "Value," and others "Economy." Review of this research raised the question of whether similar segments existed for health insurance.

Our finding of two distinct and definable segments in the health care market has recently been supported by an article published in American Demographics. In April, 1989, Thomas and Sehnert reported on "The Dual Health-Care Market." In particular, the authors identified "Traditional" and "Modern" markets -- analogous to the "Traditional" and "HMO" markets we identified in the Iowa marketplace.

Data Set Characteristics

Five different studies were conducted by BCBSI over the past four years: four conjoint studies and one perceptual mapping study. Two of the studies were conducted with random samples of Iowans; the remainder with client groups. Exhibit 1 illustrates the basic information about the sample composition and survey methodologies used in these studies. Each study was designed for a unique purpose, and the attributes were custom designed for the types of insurance products under consideration. While there are overlaps in the attributes and particular levels included in the studies, the similarities that we will note in segment composition are not simply an artifact of identical studies being conducted with very similar consumers.

Exhibit 1

Blue Cross and Blue Shield of Iowa
Research Included in Segmentation Study

<u>Service Unit Study</u>

Sample: 201 adult Iowans covered by health insurance

Methodology: Adaptive Conjoint Analysis

Attributes: Service components of health insurance policies

<u>Image Study</u>

Sample: 403 adult Iowans covered by health insurance

Methodology: Adaptive Perceptual Mapping

Attributes: Image oriented, with some product characteristics

<u>Client Study 1</u>

Sample: 119 employees of a mid-Iowa retailer

Methodology: Adaptive Conjoint Analysis

Attributes: Product and service characteristics of health insurance plans

<u>Client Study 2</u>

Sample: 88 employees of an eastern Iowa county

Methodology: Adaptive Conjoint Analysis

Attributes: Product and service characteristics of health insurance plans

<u>Client Study 3</u>

Sample: 400 employees of a central Iowa company

Methodology: Adaptive Conjoint Analysis

Attributes: Product and service characteristics of health insurance plans

METHODOLOGY

Two different clustering methodologies were used to analyze the underlying structure of the five data bases. The first one described uses the statistical procedures included in SPSS/PC+ to develop variable clusters and attendant respondent clusters. This multi-stage process begins with clustering of attribute levels and uses these attribute level clusters to define centroids that enable clusters of individuals to be formed and profiled. The second uses the CCA software developed by Sawtooth Software that clusters respondents based on full attribute-level response profiles.

Attribute Based Clustering

The cluster analyses conducted using this procedure were conducted on the five data sets before CCA was commercially available. The methodology developed is a compromise design that allowed clustering of large data bases using SPSS/PC+ QUICK CLUSTER. The analysis steps used are as follows:

1. Utility values for individual attribute levels were examined to determine which were most important to the consumers included in a particular study. The most preferred level from every attribute was always included. Unimportant attribute levels were excluded from the clustering analysis.

2. The remaining attribute levels were clustered as variables using the CLUSTER routine included in SPSS/PC+, and the resulting cluster structure was examined for its plausibility and internal consistency. These variable clusters formed the basis of the consumer segments that were identified and profiled in succeeding steps.

3. CLUSTER analysis was repeated with subsets of the attribute levels until a reasonable and believable solution was found. Attribute levels that clustered "too closely" with another level of that same attribute were excluded at each step.

4. Once an understandable and believable cluster of attribute levels was reached, these clusters were used to define centroids for clustering individual respondents using QUICK CLUSTER. Centroids were initially defined using "above-average" utility values for the attribute levels included in a cluster solution and average levels for all other attribute levels. Membership of individuals in these clusters was determined through the QUICK CLUSTER algorithm, and consumer profiles were prepared.

Respondent Based Clusters

CCA was used to cluster respondents into market segments. To implement the CCA methodology, we used the suggestions offered by Sawtooth Software in the User's Manual. The data in the UTIL.PTS files were standardized and centered, and a mixed starting point strategy was utilized. No "hold-out" sample was used, as our approach to this research was exploratory in nature. The average reproducibility and pooled-F statistics were used to diagnose the results and select the best clustering solution. The tables of Reproducibility Due to Chance Alone in Appendix E of the CCA Users' Guide were used to test the null hypothesis that definable clusters do not exist in these data sets.

CLUSTER ANALYSIS FINDINGS

The Service Study

This study was designed to isolate the "service" components of health insurance products -- for example, claims processing time, personality of customer service reps, and claims filing requirements -- from the "insurance" aspects. The objective behind applying cluster analysis techniques to this data base was to determine whether different groups of consumers respond differently to varying service packages. If this was true, "service bundling" could be used to attract consumers to health insurance options that are appropriate for them from an actuarial perspective.

The service study utilized conjoint analysis and included such attributes as "how easy it is to understand the policy," "how do insurance representatives treat you," "how frequently are claims updates mailed," and "is the claim paid directly to the physician or not." Respondents were also asked whether they would be willing to pay a higher monthly premium for more preferred service packages. Not surprisingly, consumers as a whole prefer simple and understandable coverage, helpful and informed insurance reps, and "one contact" resolution of claims questions. Consumers do not want to pay more for enhanced service packages, and having the doctor paid directly is an important convenience. Less emphasis is placed on how communication with the insurance company occurs.

Variable Based Clusters

Clustering of variables results in three segments in this service study -- the "Efficients," "Automatics," and the "Personals." The "Efficients" are the smallest segment in this sample, and prefer a policy that is easy to understand, with one contact to the head office to resolve claims, and clarity of communications.

The largest segment, the "Automatics," desire very little interaction with the insurance company and are willing to pay more for convenience. They do not want to file their own claims, expect quick payment, and would like a contact at their company who is knowledgeable about insurance.

The remaining respondents desire the personal touch in insurance products -- thus the name, "Personals." These respondents are willing to visit the insurance company personally to resolve claims and want updates on the status of claims and deductibles. They would like a personal insurance representative. The "Personals" are more likely to buy their own insurance and have more experience with filing claims.

Market share simulations for these respondents support the finding that "more is better." Many respondents prefer the service advantages and convenience that are offered by HMO-type insurance coverage.

Table 1

Market Share Simulations

Cluster	Commercials	HMOs
Efficients	53%	46%
Automatics	55	45
Personals	55	46

Respondent Based Clusters

CCA clusters of this data set revealed two segments of respondents with differing needs and desires for service bundles. The larger cluster values clarity and consistency in the services offered by insurers. These individuals value "knowing the rules" associated with resolving questions, and desire a system where claims questions are resolved through the mail via the main office. They want to know where they stand on a regular basis and are willing to pay more to obtain this kind of system. The response patterns of these individuals reveal that these customers presume that certain things will happen "automatically" in the system, such as claims being paid in ten days or payment will be made directly to the physician. It is possible that these individuals have not experienced any difficulties in resolving claims, or that they have limited claims-filing experience.

The other cluster wants a simple system for interacting with the insurance carrier. They want a system where they have an individual to contact to resolve claims questions. They expect claims to be paid directly to the physician within ten days of filing, and they do not want to file claims themselves. They are not interested in a system that involves indirect contact or increased costs.

Market share simulations were prepared for each of the two segments. The unusual finding is that all respondents prefer the high levels of service offered by HMOs, although the "HMO Segment" exhibits a stronger preference than the "Traditional Segment." The remaining shares of preference are evenly split between service packages typical of BCBSI plans and those of other commercial carriers.

Table 2

Market Share Simulations

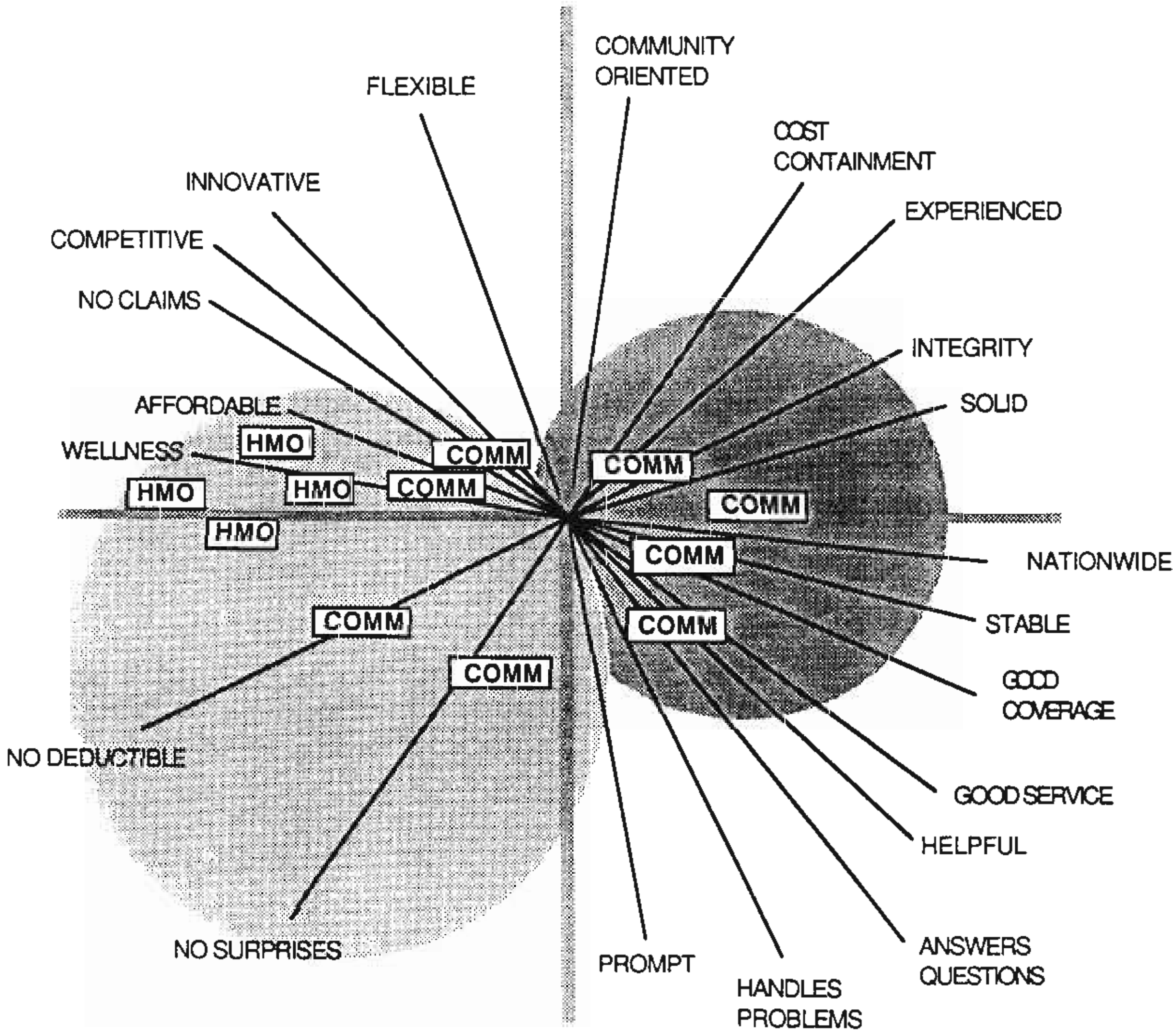
Cluster	Commercials	HMOs
Traditionals	56%	44%
HMO	47	53

The Image Study

This study was one of the earliest research efforts undertaken by Blue Cross and Blue Shield of Iowa. The objective was to understand the perceptions that Iowans have of BCBSI in the marketplace, compared to our many competitors. Perceptual mapping was used to develop visual images of the health insurance market for Iowa, and to show the perceived position of various companies within this market. The basic map of the marketplace is illustrated in Exhibit 2. Of note is the fact that the HMOs cluster (in terms of customer perceptions) along the left edge of the horizontal axis. The other commercial carriers included in the research are located centrally and to the right. It is clear from this visual presentation that consumers distinguish among the various health care alternatives available to them in the marketplace. The question is whether there are segments within this market with differing perceptions and accompanying product preferences. Clustering was done on respondent's perceptions of health care alternatives.

Exhibit 2

CONSUMER MARKET STRUCTURE PERCEPTIONS



Variable Based Clusters

The clustering of variables identified three clusters in this perceptual mapping study. To form these clusters, only service and product attributes were included. "Imagery" attributes were excluded, as the primary interest was focused toward service and product design. Again, these clusters could be described as "Efficients," "Automatics," and "Personals."

The "Efficients" are the largest group in this segmentation strategy. These respondents want an insurance company that is helpful and trustworthy, pays claims promptly, handles problems fairly, and answers questions quickly. They also want the insurance company to provide good, accurate service.

Those consumers desiring "Personal" service want their insurance company to be friendly and caring and flexible in dealing with claims problems or questions. They do not want any surprises from the insurance company. These respondents represent the smallest group in the sample and are less likely to receive insurance coverage from their employers. The last group, the "Automatics," represent the remainder of the sample, and are only concerned about one characteristic of insurance policies: coverage where there are "no claims forms to send in."

Respondent Based Clusters

When CCA was applied to this data, two stable clusters of respondents were identified. The smaller cluster includes individuals who want a traditional insurance carrier -- much like Blue Cross and Blue Shield of Iowa or the other commercial carriers. These respondents are older, with lower incomes than other who participated in this research. Demographically, these respondents are similar to those described by Thomas and Sehnert. The perceptual map for these respondents (Exhibit 3) illustrates their view of the insurance marketplace. The horizontal dimension represents the traditional characteristics of insurance. The vertical dimension represents "service and human" characteristics.

The second cluster includes respondents who prefer the convenience of a "non-traditional" approach to health care. Their attributes of importance: no claims forms to send in, no annual deductible, community oriented, innovative, and non-profit. The horizontal dimension of the perceptual map for these respondents (Exhibit 4) is characterized by typical HMO characteristics. The vertical dimension represents traditional insurance.

Exhibit 3

TRADITIONAL MARKET PERCEPTIONS

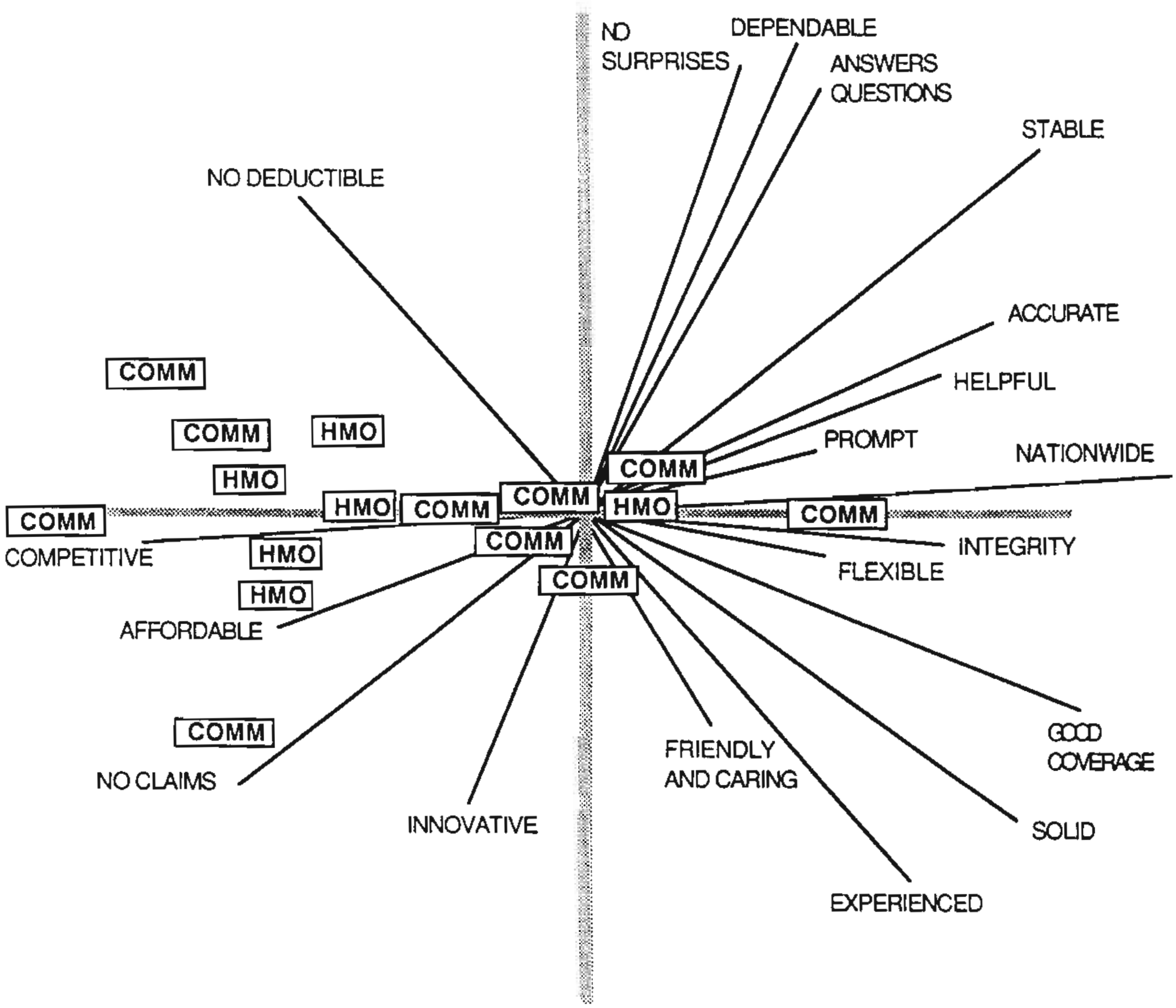
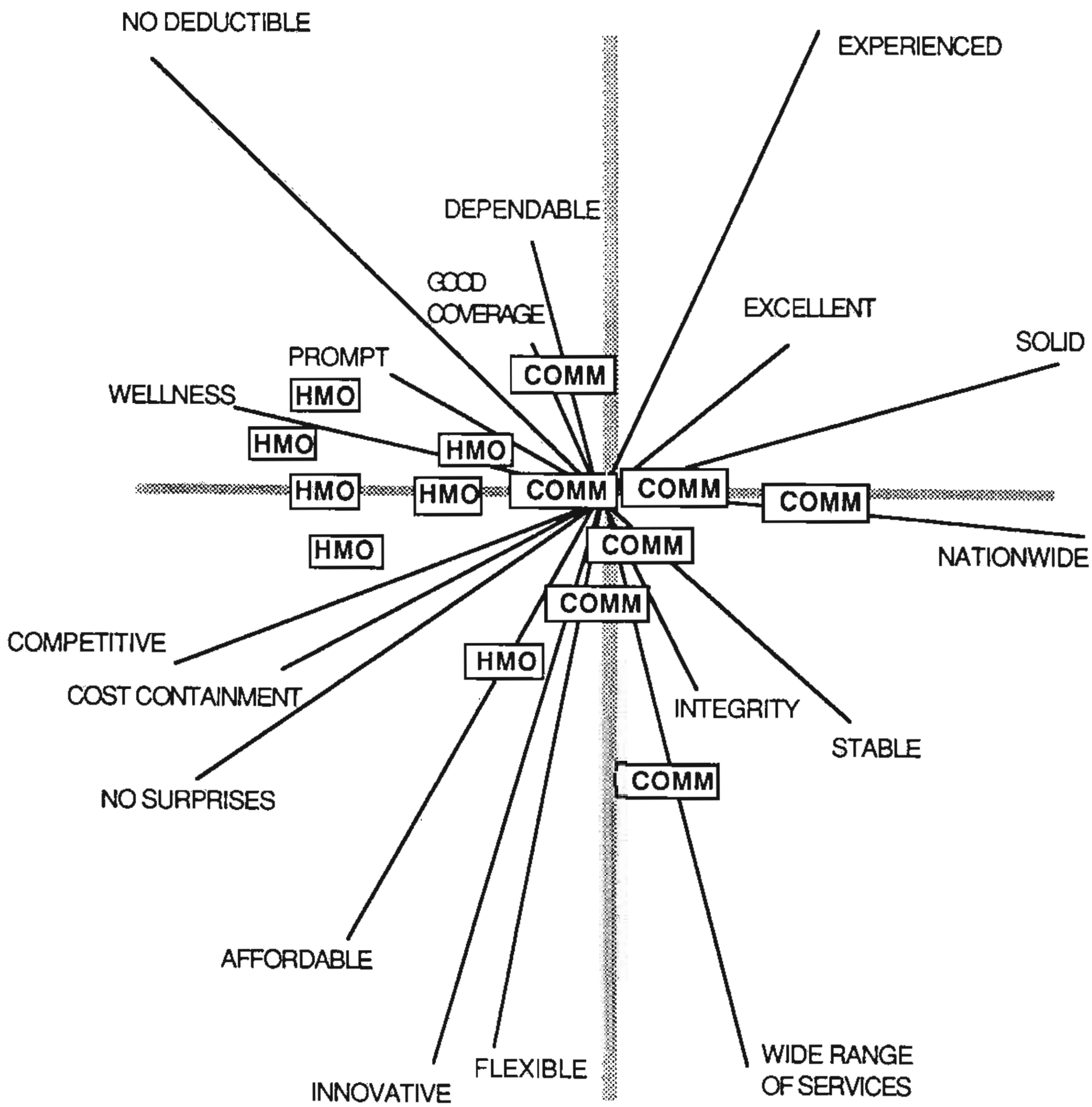


Exhibit 4

HMO MARKET STRUCTURE PERCEPTIONS



The Client Studies

The three client studies were designed to explore the preferences of consumers who are faced with multiple health insurance alternatives. The respondents in all three studies work for companies that either have implemented or will be implementing insurance options where the employee has a choice from among several insurance alternatives. Typically, these choices consist of traditional comprehensive major medical (CMM) insurance programs, managed care options like preferred provider organizations (PPO,) and health maintenance organizations (HMO).

Of interest to insurers is how consumers will make their product selection: Will the final selections of consumers be based on level of coverage, style of health care coverage, or anticipated patterns of health care utilization? And are there segments of consumers with clearly definable decision styles who will be attracted to particular products? Although we typically use these studies in a consulting role to help employers understand the needs and desires of their employees, the information gathered from these studies is very interesting to BCBSI from a product development and design standpoint.

These three studies are conjoint studies where the attributes are defined as components (both service and insurance features) of health insurance alternatives that the company is considering making available to their employees. The attributes can be bundled to create product configurations that match the products currently offered in the marketplace. The respondents in two of these studies have not experienced a choice of health insurance options before, and are therefore naive about implications of the choices they might make. The third group of respondents have experienced making choices, and should understand the implications of their choices.

Reanalysis of all three data sets indicates the presence of two or three segments in the data. One of the segments in each case is an "HMO" segment -- appreciative of the convenience and full coverage of HMOs. The second segment is a "Free Choice" segment that prefers to have no limitation on the choice of physician. These respondents prefer a more traditional approach to the provision of health insurance. The third segment that exists in the data bases is harder to explain and understand. We hypothesize that this segment may consist of individuals who cannot integrate the complex information presented to them in a multiple-choice environment, and whose pattern of product and service preferences is therefore not as clear as for the other two segments.

In this report, the results for only one of the three data sets will be reported. All three data sets lead to similar conclusions, and the patterns of preferences and market shares are likewise similar across the data. The data that will be reported are for the 400 respondents of a mid-Iowa corporation. These respondents have made health insurance choices in the past, and are familiar with the differences in coverage offered by the alternatives available to them.

Variable Based Clusters

Two strong clusters were identified from the variable based clustering methodology. The "HMO" cluster accounts for the largest proportion of the respondents. The preferences of this cluster are typical of the coverages offered by HMO products: All charges for all services are covered for participating physicians and pre-approval is necessary for some procedures and hospitalizations.

The other cluster values the freedom to choose their own physicians. This "Free Choice" cluster tends to be better educated and married. The product preferences exhibited by this group include the ability to choose any doctor, no annual out-of-pocket expenses, nationwide coverage, and no pre-approval requirements for treatment.

Two sets of market share simulations were run for each segment. The first simulation includes all of the service and product attributes included in the study. The second simulation excludes the service attributes. With all attributes included, consumers in both clusters prefer HMO alternatives. When the service characteristics are excluded, however, there are significant preference shifts toward the CMM and PPO care alternatives. Tables 3 and 4 illustrate these preferences.

Table 3

Market Share Simulations

All Attributes Included

Cluster	CMM	PPO	HMO
HMO	17%	15%	67%
Free Choice	23	32	45

Table 4

Market Share Simulations

Service Attributes Excluded

Cluster	CMM	PPO	HMO
HMO	19%	20%	62%
Free Choice	23	38	38

Respondent-Based Clusters

CCA again identified two main market segments: the "Free Choice Segment" and the "HMO" segment. The "Free Choice" segment includes respondents who prefer those characteristics normally associated with CMM and PPO product alternatives. These include the freedom to choose a physician, limited annual deductibles and out-of-pocket costs, and require that the consumer file all medical claims.

The "HMO" segment desires full coverage for all services, even if this means limitations on their choice of physicians. They also like automatic claims filing and the no annual out-of-pocket cost provisions of HMO coverage. These younger consumers have higher educational attainment than others included in the research.

The market simulations for these two segments mirror the pattern found with the variable based-clusters. Again consumers switch from HMO-style coverage when service attributes are excluded from the simulations.

Table 5

Market Share Simulations

All Attributes Included

Cluster	CMM	PPO	HMO
HMO	18%	13%	70%
Free Choice	22	32	46

Table 6

Market Share Simulations

Service Attributes Excluded

Cluster	CMM	PPO	HMO
HMO	19%	18%	63%
Free Choice	22	37	40

A Market in Transition

In this data set there is considerable evidence of a market in transition. Our belief is that the market is evolving, and that the impetus for this evolution comes from the consumers themselves. The development of alternative health insurance coverage systems such as health maintenance organizations (HMOs) and preferred provider organizations (PPOs) has not only revolutionized the health insurance industry, but the thinking of the consumer as well.

In this client study, two clear market segments emerged: a segment that strongly desires freedom of choice and a segment that prefers an HMO alternative. Diagnostically, however, the five-cluster variable-based cluster is as plausible as the three-cluster respondent-based cluster. While difficult to summarize, careful and repeated application of these methods leaves us with the opinion that these additional segments are not ephemeral or transitory, but are very real indicators of an evolving market.

Variable Based Clusters

The five-cluster solution here is as plausible as the strong two- cluster solution. It would be easy to dismiss this solution, except for the fact that the cluster sizes are large enough to be considered significant, and the demographic compositions of the clusters differ. It is possible that there are customers in this market who are not satisfied with the health insurance coverage decisions they have made in the past. This is best illustrated by Cluster Two, where the grouping of attribute levels does not make sense. This cluster is willing to pay $20 more a month for a program that requires pre-approval. This could indicate that they are willing to pay more while giving up some of their decision making freedom.

This combination of attributes may indicate that they believe they are not currently getting the kind of coverage they might prefer. Table 7 breaks out the attribute levels that dominate for each of these five clusters.

Table 7

Five Cluster Solution

Cluster One (19% of the sample)

- Full payment for certain doctors
- All services covered
- Deductibles for ER and physician visits

Cluster Two (21% of the sample)

- Pay $20 more per month
- Pre-approval required

Cluster Three (21% of the sample)

- Full payment for any doctor
- Pre-approval not required
- No annual out-of-pocket expenses

Cluster Four (25% of the sample)

- Claims filed automatically
- Pay $40 more per month

Cluster Five (15% of the sample)

- Pays at 80% for any doctor
- Pays at 90% for certain doctors
- Unlimited nationwide coverage
- Pre-approval sometimes required

Respondent-Based Clusters

The CCA clustering methodology identified three viable clusters. The "HMO" and "Free Choice" clusters are similar to those described before. The third cluster, which is the smallest of the three, is more difficult to interpret. This group of respondents values the following characteristics of health insurance:

Pay $40 more a month
Pre-approval always required
Limited annual deductibles
Must file all claims
Regular updates on claims status
Limited annual out-of-pocket costs

Demographically, this third cluster differs from the sample in one way. This group of respondents has significantly lower educational attainment than other respondents. It is likely that this group is trying to avoid catastrophic costs often associated with health care. Another possible explanation is that these respondents are not satisfied with health insurance decisions they have made in the past, but do not know how to integrate complex coverage information to make a better or wiser decision.

From a technical standpoint, this market segment is very stable as the number of clusters increases from three to four and more. CCA output includes a "switching matrix" that details how many respondents switch from cluster to another as the number of clusters increases. Very little "switching away" from this cluster occurs as the number of clusters increases.

CONCLUSIONS

There is considerable evidence in the recently completed segmentation work that the health insurance marketplace is in transition. The belief of the primary researchers is that the market is evolving, and that the impetus for this evolution comes from the consumers themselves. The development of alternative health insurance coverage systems such as health maintenance organizations (HMOs) and preferred provider organizations (PPOs) has not only revolutionized the health insurance industry, but the thinking of the consumer of health insurance as well.

The five data sets examined by BCBSI give clear indications of two strong consumer segments in the market: a traditional insurance segment and an HMO segment. These two segments are present and stable across the data sets. Several of the data sets, however, exhibit less stable, but potentially significant additional segments. The preferences of these segments are not always clear, but their presence cannot be denied.

The diagnostics for the client study reported indicate that the five-cluster variable solution may be worth examining. Although the two-cluster solution is the strongest, and clearly optimal, the five-cluster solution indicates the presence of two or three market niches whose product and service expectations may not be well met by either traditional or HMO products. The question is whether these consumers are slipping through a gap somewhere, and whether they might be more likely than others to change their coverage.

While difficult to summarize, careful and repeated examination of these two data sets leaves us with the opinion that the additional segments observed in the data are not ephemeral or transitory in nature, but are very real indicators of an evolving market.

The existence of distinct and definable groups of consumers with differing preferences for levels and forms of service raises an opportunity for Blue Cross and Blue Shield of Iowa. There is the potential to integrate service into future product development and research in a way that enhances the probability that individual consumers will be attracted to the product that is most appropriate for them from an actuarial perspective. This research demonstrates that bundling of product and service features can be used to develop optimal products.

Note: Part of this research was conducted on behalf of Blue Cross and Blue Shield of Iowa while the author was Senior Research Associate at IMR Systems, Ltd. in Des Moines, Iowa. The assistance of Mary Ellen Burr in programming and analyzing these data sets is gratefully acknowledged.

References

Kotler, P. and Clarke, R. (1987). Marketing for Health Care Organizations. Englewood Cliffs, N. J.: Prentice Hall, Inc.

Thomas, R. K. and Sehnert, W. F. (1989, April). "The Dual Health-Care Market." American Demographics, pp. 46-47.

Woodside, A. G., Nielsen, R. L., Walters, F. and Muller, G. D. (1988). "Preference Segmentation of Health Care Services: The Old-Fashioneds, Value Conscious, Affluents, and Professional Want-it-Alls." Journal of Health Care Marketing, 2 (2), p. 14.

A COMPARISON OF 18 CLUSTERING ALGORITHMS GENERALLY AVAILABLE TO THE MARKETING RESEARCH PROFESSIONAL

William D. Neal
SDR, Inc.

OVERVIEW

This article compares eighteen clustering programs generally available to marketing researchers. Each program was tested using three data sets - Fisher's Iris data, Fisher's Iris data with a randomized case order, and a real market segmentation database containing 994 cases of 75 measures each, using a 7-point scale. The results show predictable and rather consistent solutions on the Fisher's data sets. However, there were extremely wide variations in the solutions to the larger data set. Several measures of clustering accuracy were used - percent clustered correctly (on the Fisher's data only), percent of variance explained, a homogeneity index for each cluster, and percent agreement between the solution on the Fisher's data set, and the randomized Fisher's data set. Preliminary conclusions and limitations on the study are discussed.

INTRODUCTION

Cluster analysis is often considered a "black art" by both academic and practicing statisticians. There is no generally accepted philosophical definition for the procedure, nor is there a generally accepted statistical definition for the procedure.

Cluster analysis is extremely sensitive to levels of measurement, to data scales, and to the algorithmic approach. Variations in these factors often give considerably different results on the same data set.

Yet, cluster analysis is a useful and necessary tool of marketing research. It allows us to group like-type markets, group like-type products and services, and group like-type consumers and customers.

Cluster analysis is an essential analytic tool for market segmentation. Fundamentally and intuitively we know clusters of consumers and customers very often exist such that they react similarly to product/service attributes within clusters (or segments), and react differently to product/service attributes between clusters. Given that these differences do exist, the essential tenet of market segmentation is that one can vary the marketing mix to target selected segments, thus gaining marketing efficiency and improved sales via deeper penetration of selected markets.

PROBLEMS WITH CLUSTER ANALYSIS

Broadly speaking, there are five problem areas one must address to execute a successful cluster analysis.

1. Getting the "right" measures or variables. The variables one submits to clustering are usually called the "basis" variables. The selection of the basis variables is a general research problem and not specific to cluster analysis.

2. Defining the "right" relationship between variables. In cluster analysis, one must decide whether to adjust the basis variables for scale differences, adjust basis variables for inter-correlations, adjust basis variables for differences in respondent scale utilization (i.e. centering), and whether the basis variables should be weighted.

3. Defining the "right" relationship between groups. The analyst needs to address what will be done with outliers, what levels of density (homogeneity) are acceptable, what levels of dispersion (variance) are acceptable, and whether cluster overlap will be allowed.

4. Finding the "right" algorithm. Unfortunately, clustering algorithms vary widely on their assumptions about the form of the data. Some assume that the data are spherical in multidimensional space (i.e. hyper-spherical), and therefore tend to find solutions under that assumption. Others assume there are chains, or natural linkages, or a hierarchy between observations in multidimensional space.

5. Knowing when you have the "right" number of clusters. Although there have been many attempts to statistically determine the optimum number of clusters in a cluster analysis, none have gained general acceptance and none seem to be applicable to every marketing research clustering problem. Ultimately, selecting the right number of clusters is still a judgment call on the part of the analyst.

THE DATA FOR CLUSTER ANALYSIS

The initial data set for cluster analysis in most marketing research applications is an N by P matrix of N cases drawn from a population of interest and measured on P variables, which we select in hopes they represent an appropriate basis for clustering. Often these P variables are rating scales which we assume are interval-level measures.

USUAL OPTIONS FOR CONDITIONING THE DATA

The next step is to determine whether the data need any type of transformation or conditioning. There are several options.

1. One could use the raw data without any transformations. In this case each variable is considered to have equal weight in defining the clusters. If the data are actually inter-correlated then this has the effect of weighting the clustering solution toward those sets of variables that are most highly correlated with each other.

2. If the basis variables are measured on different scales - for example, age measured in years, consumption measured in ounces, and income measured in thousands of dollars - then the larger scales will have the effect of weighting or dominating the cluster solution. To overcome this situation, if desired, one would standardize the variables so that each has a mean of 0, and a standard deviation of 1.

3. If the analyst believes that respondents used the scales differently, that is to say, in a relative manner, then it may be appropriate to "center" the data such that the average response of each respondent is set to 0.

4. To alleviate the problem of intercorrelation between variables, one could conduct a principal components analysis of the basis variables and use the principal component scores for each case as substitute variables in the clustering. Each principal component score then has equal weight and the principal components are orthogonal, or uncorrelated.

5. As an alternative, the P variables could be subjected to some other reduced space analysis, such as Guttman-Lingoes analysis, KYST decomposition, or correspondence analysis. Then one would use respondent loadings on the extracted dimensions as substitutes for the original basis variables.

One of the problems with these transformations is that clustering algorithms are extremely sensitive to them. Consequently, there will often be radical differences between the solution using raw data when compared to the solution using transformed data. These variations in results seem to be a direct function of the degree of overlap and the shape of the groups in hyper-space. Table 1 shows a crosstabulation of the cluster solution of non-standardized data compared to standardized data, using a Howard-Harris clustering algorithm on Fisher's Iris data set, which is basically spherical.

TABLE 1

		Solution for non-standardized Data			
		Group 1	Group 2	Group 3	Total
Solution for standardized data	Group 1	50			50
	Group 2		52	1	53
	Group 3		10	37	47
	Total	50	62	38	150

In this case, where the true number of cases in each group is 50, each solution is acceptable because the data in the Fisher's data base are fairly well separated, but the standardized data give a more accurate result.

As a matter of practice, it is almost always appropriate to standardize the variables prior to subjecting them to a clustering procedure, unless there is a compelling reason for not doing so. A compelling reason might include cases where volumetric data is being used, and the purpose of clustering is to form groups differentiated by volume usage profiles. Also, it is not appropriate to standardize conjoint utilities prior to using them in clustering.

When the basis variables are rating, Likkert-type, or other typical scales used in marketing research, it is usually appropriate to conduct a principal components analysis of the data set. If over 60% of the total variance is explained by the significant principal components, then one should consider using the principal component scores for each case as the basis variables for clustering. If the principal components analysis explains 40% or less of the total variance, it would be more appropriate to use the standardized (or raw) data. If the principal components solution explains between 40% and 60%, it's basically a judgment call.

UNDERSTANDING THE PROBLEMS WITH CLUSTERING ALGORITHMS

If we could peer into hyper-space where each variable represented a dimension orthogonal to all other variables or dimensions and actually see the location of each case, then most of us could intuitively group cases together, typically based on the following criteria.

1. We would look for swarms, or clouds of cases that represented locations of high density.

2. We would observe distances between swarms.

3. We would note the relative sizes of the swarms.

4. We would note the relative structure and form of the swarms - i.e., whether they tended to be rather spherical or cylindrical.

5. We would tend to ignore isolated cases, or outliers.

6. And intuitively, we would isolate and describe the various swarms that we observed.

Unfortunately, we do not have a statistical clustering algorithm that does all of this for us. In fact, most of the algorithms generally available to the marketing research community perform only task 1 and/or task 2 above.

As a demonstration in two dimensions, note figures 1 through 3.

Figure 1 represents three highly separated swarms of data measured on two variables. The swarms tend to be spherical, dense, of nearly equal size, and there are no outliers. Just about any clustering algorithm will provide a very stable consistent solution to this data set.

Figure 2 shows three swarms similar to figure 1 except that there is considerable overlap. All clustering algorithms would have a difficult time uncovering the three real swarms here. Many algorithms would find no acceptable solution, and other algorithms might indicate a two-group solution was best. The problem is that all algorithms will provide a solution for as many groups as we request. We have no way of directly knowing from the algorithm which is correct, or even optimum.

Figure 3 represents a situation where we would intuitively identify three groups even though one of the groups represents a linear interaction effect between the two variables. Unfortunately, most of the algorithms available to us would have a very difficult time finding those three solutions.

TYPES OF CLUSTERING ALGORITHMS

Brian Everitt, in his book Cluster Analysis (Second Edition), classified clustering algorithms into five types - Hierarchical Procedures (including agglomerative methods, divisive methods, and association analysis), Optimization Techniques (including K-means and hill-climbing procedures), Density Search Methods, Clumping Techniques, and Others. In reality, almost all of the cluster analysis computer programs available to the practicing marketing research community use either hierarchical agglomerative methods or K-means methods.

Hierarchical agglomerative methods sequentially link individual cases, or other clusters, based on some measure of distance between entities. Typically, each case starts out as a unique cluster, then the cases are successively combined to others in a growing hierarchy. Most of these methods are highly dependent upon, and sensitive to, the initial start points and initial aggregation of cases into a cluster. These procedures can be further classified as follows, based on the distance measure used.

* Single Linkage - AKA nearest neighbor, minimum method, connectiveness method.
* Complete Linkage - AKA Farthest neighbor, maximum method, diameter method.
* Average Linkage.
* Centroid Method.
* Median Method.
* Ward's Method (which can also be considered an optimization technique).

Optimization techniques usually partition a data set into successively smaller clusters. Typically, they start by locating temporary centroids for each cluster, called "seeds," and calculate initial variances or squared distances about those seeds. Some of these algorithms then iterate, reclassifying each respondent into another cluster, attempting to minimize the within group variances or distances. Either after or during each iteration of moving the cases from cluster to cluster, the various cluster centroids (or K number of cluster means) are adjusted, based on the addition or loss of cases from the cluster. The variances or distances within each cluster are recalculated, and the iteration of cases starts over. This process is continued until the centroids stabilize. Those methods that do not go through this iteration procedure (such as SPSS Quick Cluster and SAS FASTCLUS) are extremely sensitive to the location of the initial seeds.

Optimization methods do not assume any hierarchical relationships among entities or groups. They do assume that the data is somewhat spherical in hyper-space.

There are many algorithms available for this general procedure, and they vary widely in their specific approaches. Some of the more popular and more available ones include:

* Howard-Harris procedures, or Carmone's algorithm
* Convergent Cluster Analysis (CCA by Sawtooth Software)
* BCTRY
* Q.Cluster (P-Stat, not to be confused with Q-type factor analysis)
* Quick Cluster (SPSS)
* FASTCLUS (SAS)
* K-Means (MacQueen)
* Leader (Hartigan)
* K-Means (Anderberg)
* KMEANS (SYSTAT)
* Hill-Climbing

Returning to our analysis of the data distributions, figure 4 shows a possible placement of initial start points, or seeds, in the same data set shown in figure 3.

Given this distribution of swarms, a hierarchical procedure may be able to find the correct solution as shown in figure 5 because it has the ability to chain together cases in a hierarchical manner.

An optimizing procedure is more likely to give a solution as shown in figure 6 because it tends to find spherical solutions in multidimensional space.

On the other hand, a hierarchical procedure may provide radically incorrect solutions as depicted in figures 8 and 9, whereas the iterative optimizing procedures would continue to provide a solution similar to figure 6.

DATA USED IN THE COMPARATIVE TEST

Three data sets were used in the comparative test of the 18 different algorithms.

1. The Fisher's Iris data set contains a total of 150 cases, 50 for each of three different types of Iris flower. There are four variables or measures for each case - petal width, petal length, septal length, and septal width - all measured in millimeters. The data are metric and the data distributions tend to be spherical. See figure 10 which shows a scatter diagram of septal length plotted against petal length. Group 1 tends to be highly separated from groups 2 and 3, which overlap somewhat.

2. A second data set from Fisher's Iris data was formed exactly like the first, except the order of cases was randomized. This was done to test whether any of the algorithms were sensitive to the order of cases.

3. A disguised national probability sample data set was obtained containing 994 cases of 7-point agree-disagree scales. This data set was chosen because it is very typical of what would be obtained in a market segmentation study. Indeed, that was the purpose of this data set in the first place.

THE ALGORITHMS USED IN THE COMPARATIVE TEST

The following eighteen algorithms were used in this comparative test.

1. Convergent Cluster Analysis (CCA), a unique K-Means clustering procedure developed by Richard Johnson and published by Sawtooth Software. The processing on the three data sets was conducted by the staff at Sawtooth Software.

2. Howard-Harris K-Means. This is the original algorithm programmed by Frank Carmone and modified by SDR, Inc. The program was run on a DEC VAX 11/750 by SDR staff.

3. Howard-Harris K-Means. This is the version found in the PC-MDS package published by Scott Smith from Brigham Young University. The program was executed by SDR personnel using an IBM PC/XT.

4. Ward's Method. Three versions of this program were tested, one from PC-MDS, one from SAS, and one from SPSS. SDR personnel executed the PC-MDS and SPSS versions. The SAS version was executed by Jim Heiborn, Associate Senior Marketing Research Analyst at Hallmark using an IBM mainframe.

5. KMEANS published by SYSTAT. The program was executed by J. Walker Smith at DowBrands using an IBM-AT.

6. Q.Cluster published by P-Stat. The program was executed by the technical staff at P-Stat.

7. Three other programs from the SAS package were executed by Jim Heiborn at Hallmark. They were:

   * Single Linkage
   * Average Linkage
   * FASTCLUS

8. A number of available clustering routines were executed with SPSS-X by the SDR staff using a DEC VAX 11/750 minicomputer. They were:

   * Quick Cluster
   * Quick Cluster with iterative seeding
   * Single linkage (between groups option)
   * Single Linkage (within groups option)
   * Complete Linkage
   * Average Linkage
   * Median Method

THE CRITERION

Four criteria were used to compare the clustering results:

1. The percent of Fisher's Iris data classified correctly by the three-group solution of each algorithm.

2. The percent of variance explained by clustering. This was calculated by first calculating the total sums of squares within the data set before clustering, then calculating the sums of squares within each cluster, summed across all clusters. Percent of variance explained was calculated as follows:

$$SS_{Total} = SS_{Within} + SS_{Between}$$

$$SS_{Between} = SS_{Total} - SS_{within}$$

Converting each SS to variance (dividing by total N) and dividing by total variance to get a proportion of variance explained.

$$\frac{(SS_{Btwn}/N)}{(SS_{Tot}/N\ )} = \frac{((SS_{Tot}/N) - (SS_{Wi}/N))}{(SS_{Tot}/N)}$$

Which reduces to:

Pct. of variance between groups = $100 * (1 - (S^2_{Wi} / S^2_{Tot})$

Note that SS_{Wi} is the sum of squared deviations from the centroid within each group, summed across all groups.

3. The percent of agreement between the solution for the Iris data and the solution for the case-randomized Iris data.

4. The homogeneity of groups. This is based on a homogeneity index that is calculated by dividing the variance within each cluster by the total variance in the data set before clustering. It usually varies between 1 and 0, with a 1 indicating that the group is no more homogeneous than the unclustered data set, and a 0 indicating that all cases in the group have exactly the same value on every variable. It is posssible for the homogeneity index for a particular question to exceed 1, indicating that the cluster is even less homogeneous than the original data set.

TABLE I

A COMPARISON OF 18 CLUSTERING ALGORITHMS

THE RESULTS

IRIS DATA SET

RESULTS ORDERED BY PERCENT CORRECT, THEN VARIANCE EXPLAINED

PROGRAM NAME		HOMOGENEITY INDEX GP1	GP2	GP3	PERCENT CORRECT	PERCENT VAR. EXPLND	PERCENT AGREEMENT
ACTUAL DATA		.0667	.1348	.1917	100%	100%	
SPSS Centroid		.0667	.1347	.1454	90.70%	88.34%	
	S		(same)				100%
SPSS Quick Cluster		.0667	.1437	.1357	90.00%	88.38%	
(iterative seeds)	S	.0667	.1414	.1383	89.33%	88.43%	98.70%
SAS FASTCLUS		.0667	.1515	.1288	88.67%	88.18%	
	S	.0667	.1480	.1383	90.00	88.33%	98.7%
Howard-Harris (SDR)		.0667	.1414	.1383	89.33%	88.43%	
	S		(same)				100%
PC-MDS Howard-Harris		.0667	.1414	.1383	89.33%	88.43%	
	S		(same)				100%
SYSTAT KMEANS		.0667	.1414	.1383	89.33%	88.43%	
	S		(same)				100%
CCA - Sawtooth		.0667	.1414	.1383	89.33%	88.43%	
	S	.0667	.1435	.1382	88.68%	88.43%	99.33%

NOTE: The row marked "S" designates the randomized case order solution.

TABLE II

A COMPARISON OF 18 CLUSTERING ALGORITHMS

RESULTS ORDERED BY PERCENT CORRECT, THEN VARIANCE EXPLAINED (continued)

PROGRAM NAME		HOMOGENEITY INDEX GP1	GP2	GP3	PERCENT CORRECT	PERCENT VAR. EXPLND	PERCENT AGREEMENT
P-Stat Q.Cluster		.0667	.1414	.1382	89.33%	88.43%	
	S		(same)				100%
PC-MDS Ward's		.0667	.1467	.1314	89.33%	88.36%	
SPSS Ward's			(same)				
SAS Ward's			(same)				
All Ward's shuffled			(same)				100%
SPSS Average Link.		.0667	.1976	.0767	84.00%	86.86%	
(within groups)	S		(same)				100%
SPSS Complete Link		.0667	.1976	.0767	84.0%	86.86%	
	S		(same)				100%
SPSS Quick Cluster		.0667	.1198	.1769	82.00%	86.86%	
(random seeds)	S		(same)				100%
SPSS Median Method		.0667	.2078	.0713	81.3%	86.11%	
	S	.0667	.2103	.1053	75.3%	84.67%	92.6%
SAS Average Link		.0667	.2146	.0858	74.67%	84.50%	
(between groups)	S		(same)				100%
SPSS Single Link		.0667	.2858	.0094	68.00%	79.09%	
	S		(same)				100%

TABLE III

A COMPARISON OF 18 CLUSTERING ALGORITHMS

SEGMENTATION DATA BASE - 7-POINT SCALES, 75 VARIABLES

RESULTS ORDERED BY VARIANCE EXPLAINED

PROGRAM NAME	HOMOGENEITY INDEX GP1	GP2	GP3	GP4	GP5	GP6	PERCENT VAR EXPL.
Howard-Harris (SDR)	.8037	.7392	.7681	.6784	.7710	.7809	24.26%
CCA - Sawtooth	.8118	.8056	.7426	.7619	.6792	.0000	23.44%
SPSS Average Link. (within groups)	.7826	.7786	.7620	.7474	.7757	.7449	23.06%
P-Stat Q.Cluster	.6818	.8073	.7675	.8452	.7795	.8429	22.91%
Ward's (SPSS & SAS)	.7984	.8167	.7891	.7311	.6139	.7166	22.43%
SAS FASTCLUS	.8026	.8168	.7390	.6232	.8507	.8667	22.07%
SPSS Complete Link	.8599	.7662	.7714	1.0768	1.1413	.9337	18.96%
SPSS Quick Cluster (random seeds)	.7523	.8465	.8155	.8233	.9147	.9295	17.05%
SPSS Average Link (between groups)	.9925	.0000	.0000	.0000	.0000	.0000	1.31%
SPSS Single Link.	.9935	.0000	.0000	.0000	.0000	.0000	1.21%
SPSS Centroid	.9935	.0000	.0000	.0000	.0000	.0000	1.21%
SPSS Median	.9935	.0000	.0000	.0000	.0000	.0000	1.21%
SAS Average Link (between groups)	.9935	.0000	.0000	.0000	.0000	.0000	1.21%
SPSS Quick Cluster (iterative seeds)	NOT RUN DUE TO EXCESSIVE TIME REQUIREMENTS						
PC-MDS Ward's	WOULD NOT RUN DUE TO SIZE LIMITATIONS						
PC-MDS Howard-Harris	WOULD NOT RUN DUE TO SIZE LIMITATIONS						
SYSTAT KMEANS	WOULD NOT RUN DUE TO SIZE LIMITATIONS						

OBSERVATIONS

In reference to the comparisons using Fisher's Iris data set the following observations can be made:

1. All of the algorithms found the first group, which is highly separated from groups 2 and 3 in this data set.

2. All except the bottom three algorithms found an acceptable solution on the Fisher's data set.

3. All of the algorithms down through the Ward's procedures found extremely similar solutions.

4. All three Ward's algorithms found exactly the same solution.

5. Except for the SPSS median method, there seemed to be little sensitivity to changes in case order. This is curious, since there have been widespread allegations about the supposed sensitivity to case order by certain algorithms, especially SPSS Quick Cluster and SAS FASTCLUS.

As part of this procedure, scatter plots were generated for every possible pair of the Fisher data set variables. It appears that Fisher's Iris data set is not a very good representation of normal data sets encountered in marketing research. There are only four variables and only 150 observations. The variables are purely metric and are quite normally distributed.

In reference to the clustering of the 994 cases of 75 variables each, the following observations can be made.

1. The top six listed algorithms all produced an acceptably high percent of variance explained. Note that five out of the six algorithms are optimization procedures. It should be further noted that the CCA algorithm found a single outlier for group six. This had a significant impact on its percent of variance explained. We will address that issue in more detail further on.

2. Three programs, listed at the bottom, would not run the analysis due to their limitations on the number of variables. It is highly recommended that the publishers of these programs correct that problem.

3. Five of the algorithms initially let outliers define five out of the six groups. This is not acceptable. It is suggested that the publishers add options to throw out outliers.

4. SPSS Quick Cluster with random seeds produced two groups with disturbingly high homogeneities. These clusters would be worthless for any further analysis.

5. Based on the results using the Fisher's data set, SPSS Quick Cluster with iterative seeds would probably have performed well if there had been enough time to do this tedious task, which there was not. The publisher should program this to be an automatic option.

In order to fairly test the CCA program, the SDR Howard-Harris algorithm was rerun to produce a five group solution. The solution from this analysis was cross-tabulated against the five groups generated by CCA. The results show that the two algorithms gave very similar results. It should be noted that it is impossible to determine which algorithm is more accurate with this database because there is no "known" solution.

SAWTOOTH SOFTWARE CCA FIVE-GROUP SOLUTION

		Group 1	Group 2	Group 3	Group 4	Group 5
Howard -	Gp 1	169	0	11	0	1
Harris,	Gp 2	0	117	0	76	44
SDR version	Gp 3	0	57	127	1	3
	Gp 4	0	0	0	145	0
	Gp 5	7	33	17	0	185

In a manner similar to the above, the SPSS Average linkage solution was cross-tabulated against the Howard-Harris results. The results are somewhat comparable.

SPSS Average Linkage (within groups)

		Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
	Gp 1	1	7	79	16		58
Howard -	Gp 2		118	1	14	13	54
Harris.	Gp 3	66		94			
SDR Vers.	Gp 4		5	14	89		18
	Gp 5		14			115	4
	Gp 6		4	28	4		77

Also, the results of the P-Stat Q.Cluster algorithm were cross-tabulated against the Howard-Harris solution. The results are decidedly different.

P-Stat Q.Cluster

		Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
	Gp 1	11	6	85	3		24
Howard -	Gp 2		143				
Harris.	Gp 3		146	31			44
SDR Vers.	Gp 4	45	3	91	2	2	
	Gp 5			2	173		
	Gp 6			63	7	7	106

SPECULATIONS AND CONCLUSIONS

There are few solid conclusions that can be drawn at this time. Further investigation is required using new and different data sets. (See limitations following.) However, based on the investigation so far, the following represent a preliminary set of conclusions:

1. Although the Fisher Iris database is used by many software publishers to assess the efficiency and accuracy of their algorithms, it is not a very good database for assessing how well these algorithms will perform for the marketing researcher. In marketing research, we typically deal with a larger number of observations in the database, with more variables, and those variables are usually five-point to nine-point scales.

2. The optimization routines, including all of the partitioning routines tested, plus Ward's Method, assume that the data is hyper-spherical. Therefore it is recommended that the data be standardized (by variable) or subjected to a principal components decomposition before submitting the data to one of these clustering algorithms, UNLESS there is a compelling reason not to do so.

3. Centering respondent data (i.e. adjusting each respondent's answers as a deviation from the respondent's average response) is a philosophical decision focused on whether respondents tend to use absolute scales or relative scales. Centering will sometimes radically alter the clustering results compared to using the raw data.

4. In general, the optimization routines and Ward's Method seem to perform better in the marketing research environment than do the hierarchical distance procedures.

5. In general, it is recommended that the marketing researcher cluster the data in several ways, using both standardized and non-standardized data, using centered and non-centered data, rotating the case order, using alternative "seeds," and so forth, to see if there is a convergence of solutions. The CCA program allows the marketing

researcher to execute many of these alternatives within the program. Other software publishers ought to incorporate similar innovations.

6. It is highly recommended that the analyst conduct a large number of variable by variable scatter plots to get a feel for the data before submitting the data to clustering. The results of inspecting these scatter plots is often very revealing, and sometimes even shocking, due to the apparent lack of variability on some scaled data.

7. Ultimately, the selection of a clustering solution is judgmental, and often depends on whether the solution makes common sense.

LIMITATIONS

1. The Fisher's Iris data is not adequate for comparing clustering algorithms that will be used in marketing research. Therefore, the results based on the clustering of the Fisher data set should be suspect.

2. The national probability sample data had some missing data that was not adequately handled. In addition, there is no known, or generally accepted, "correct" solution for that data set. This severely detracted from a complete comparison of the various algorithms.

3. This study depended on several different people to actually execute the programs. There is a possibility that some program options and some program setups should have been used, but were not, and vice versa.

The author is now developing an artificial data set that will reflect the size and structure of a typical data set used in marketing research. The data set will be generated from known distributions with known characteristics. In addition, a more strict protocol is being developed to provide a more rigorous test of algorithms.

Any readers who wish to assist in this task, or who wish to execute clustering programs under this new protocol, should contact the author at SDR, Inc., 2251 Perimeter Park Drive, Atlanta, GA 30341 (404) 451/5100.

FIGURE 1

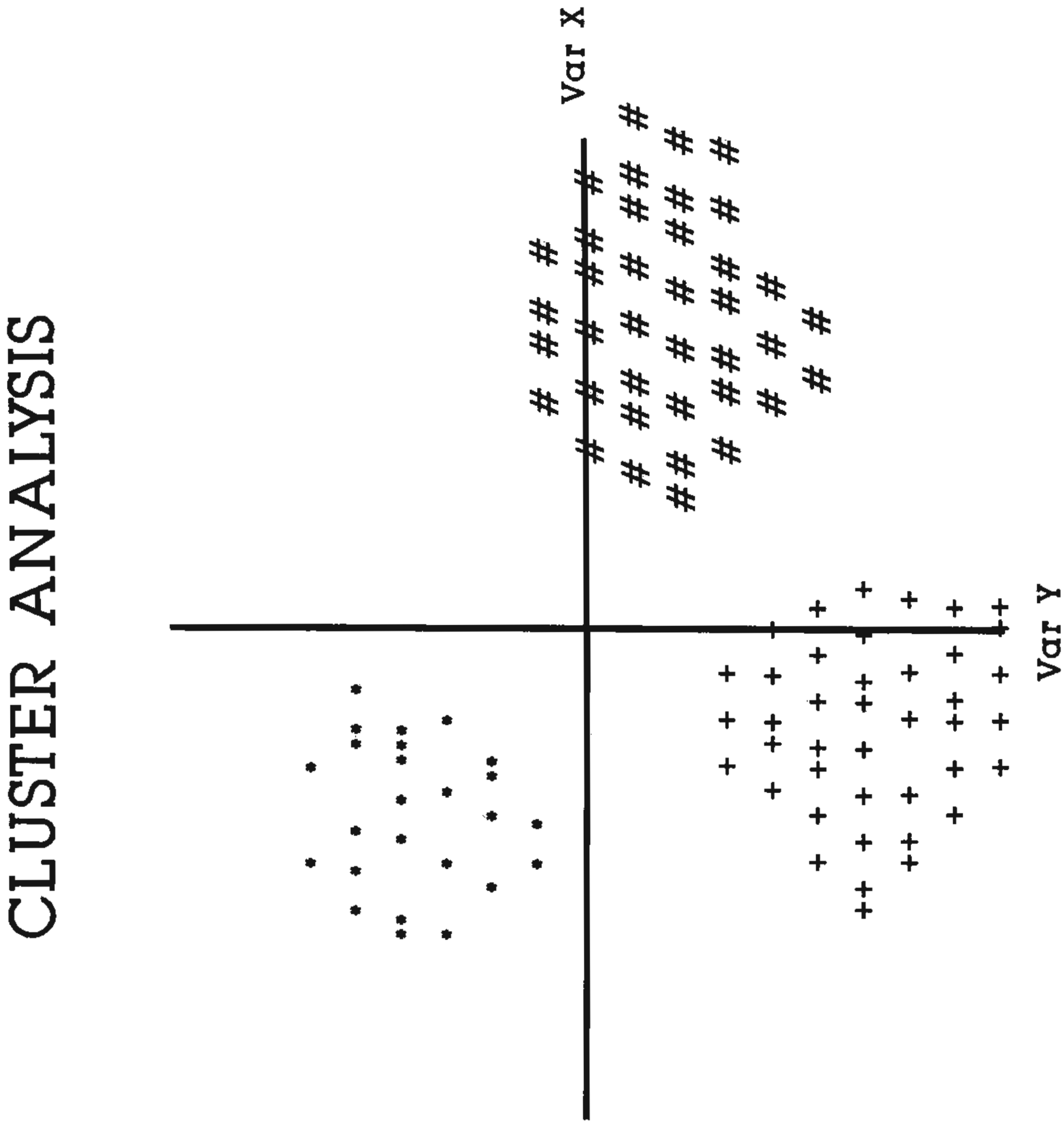


FIGURE 2

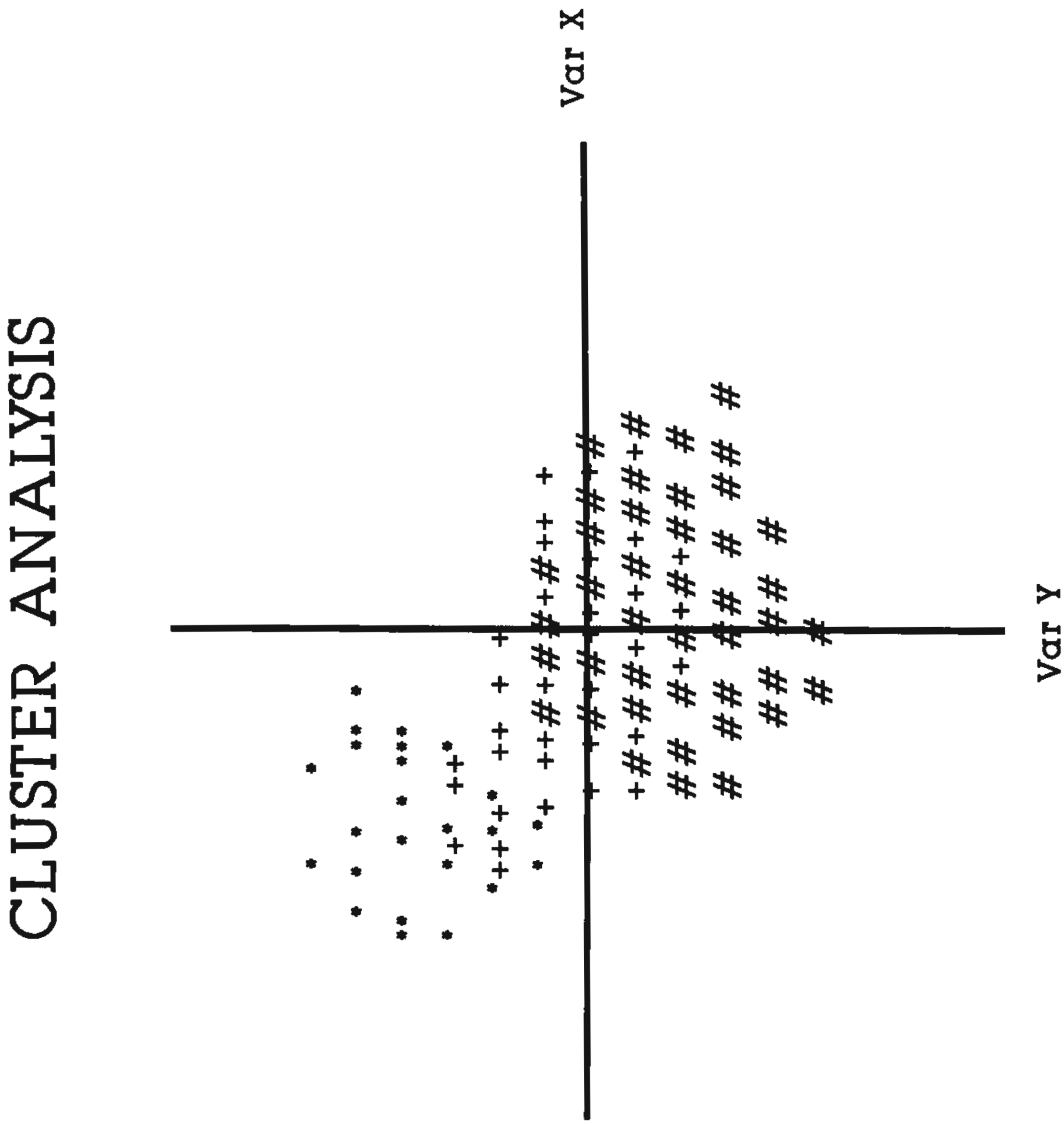


FIGURE 3

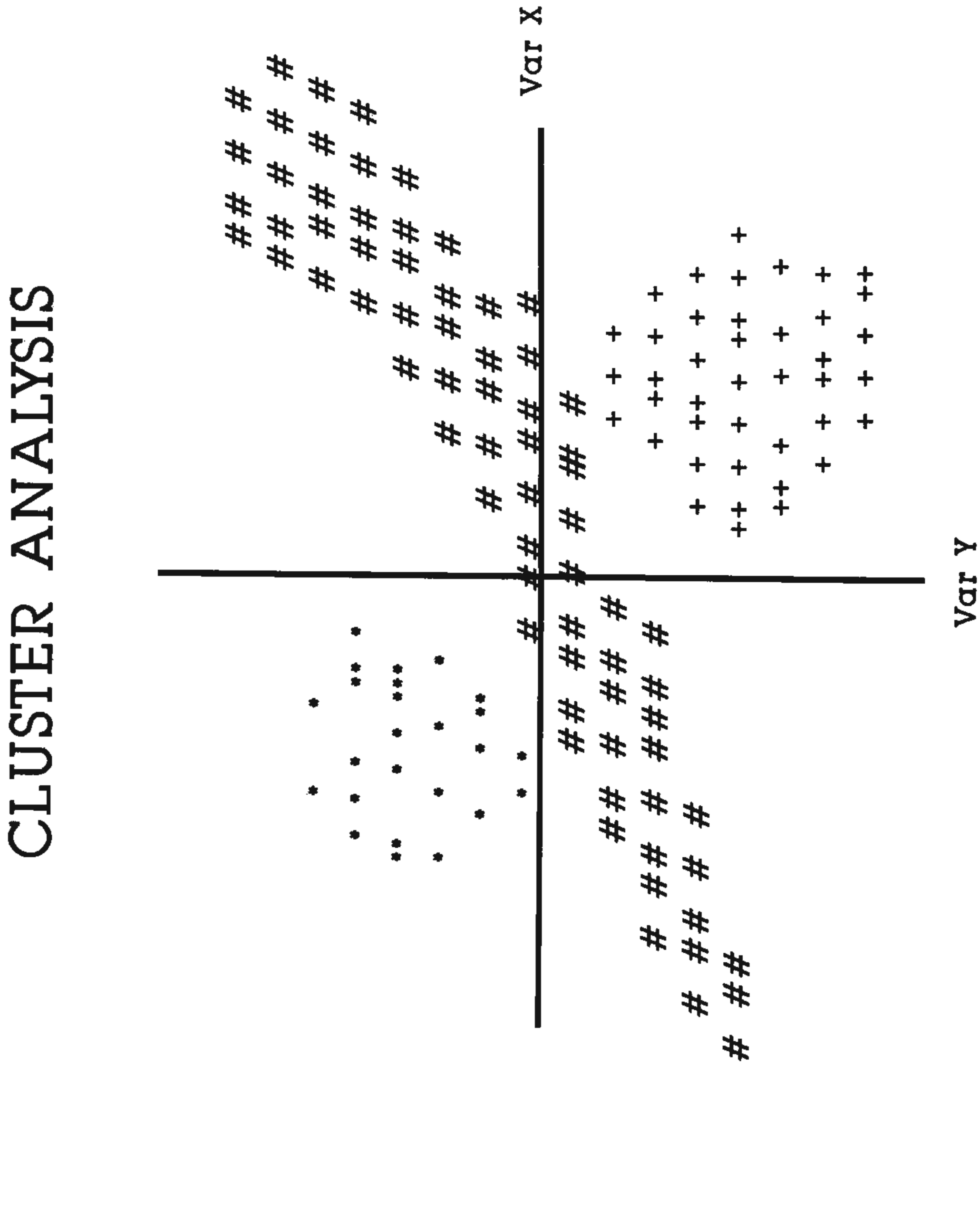


FIGURE 4

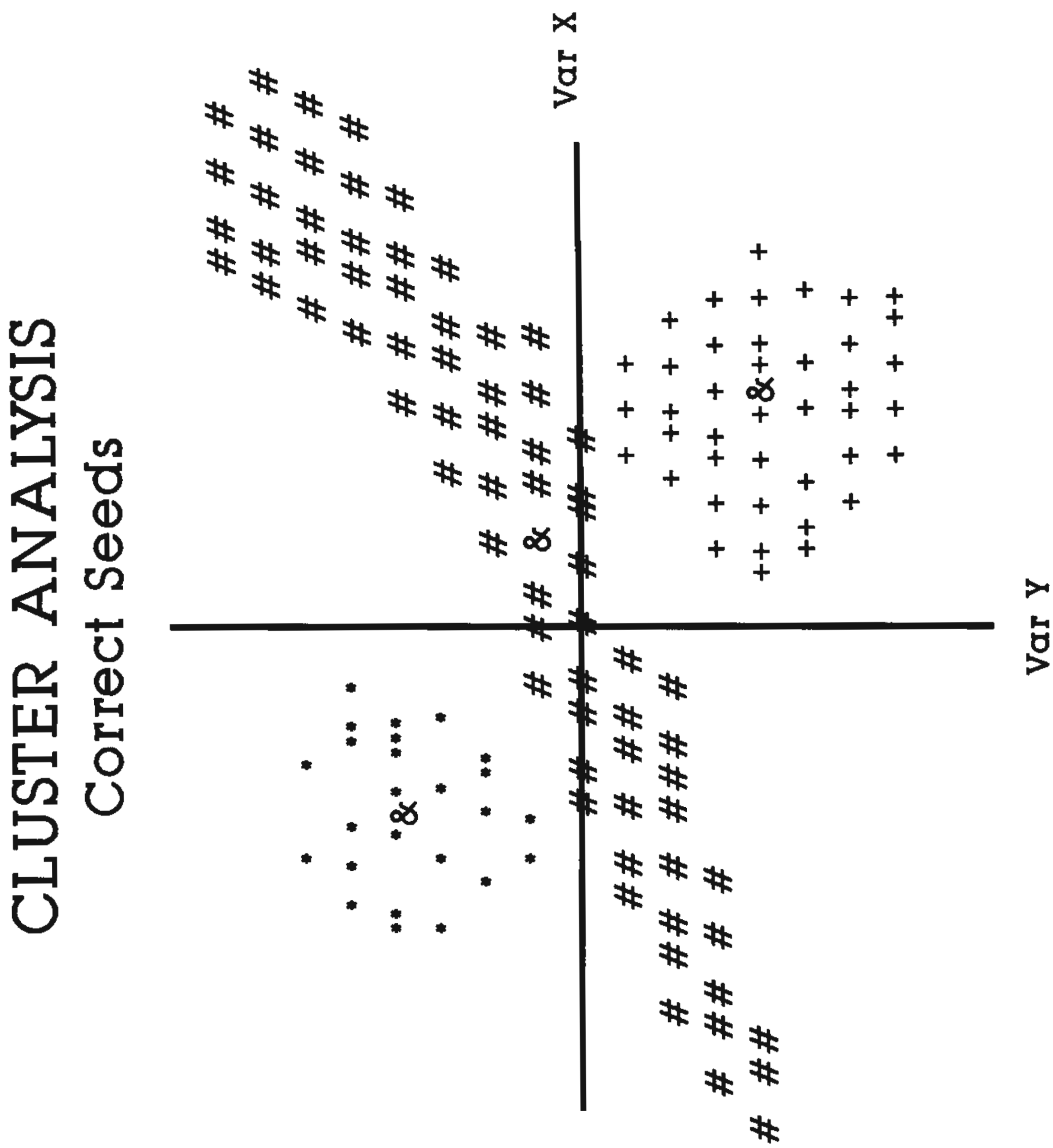


FIGURE 5

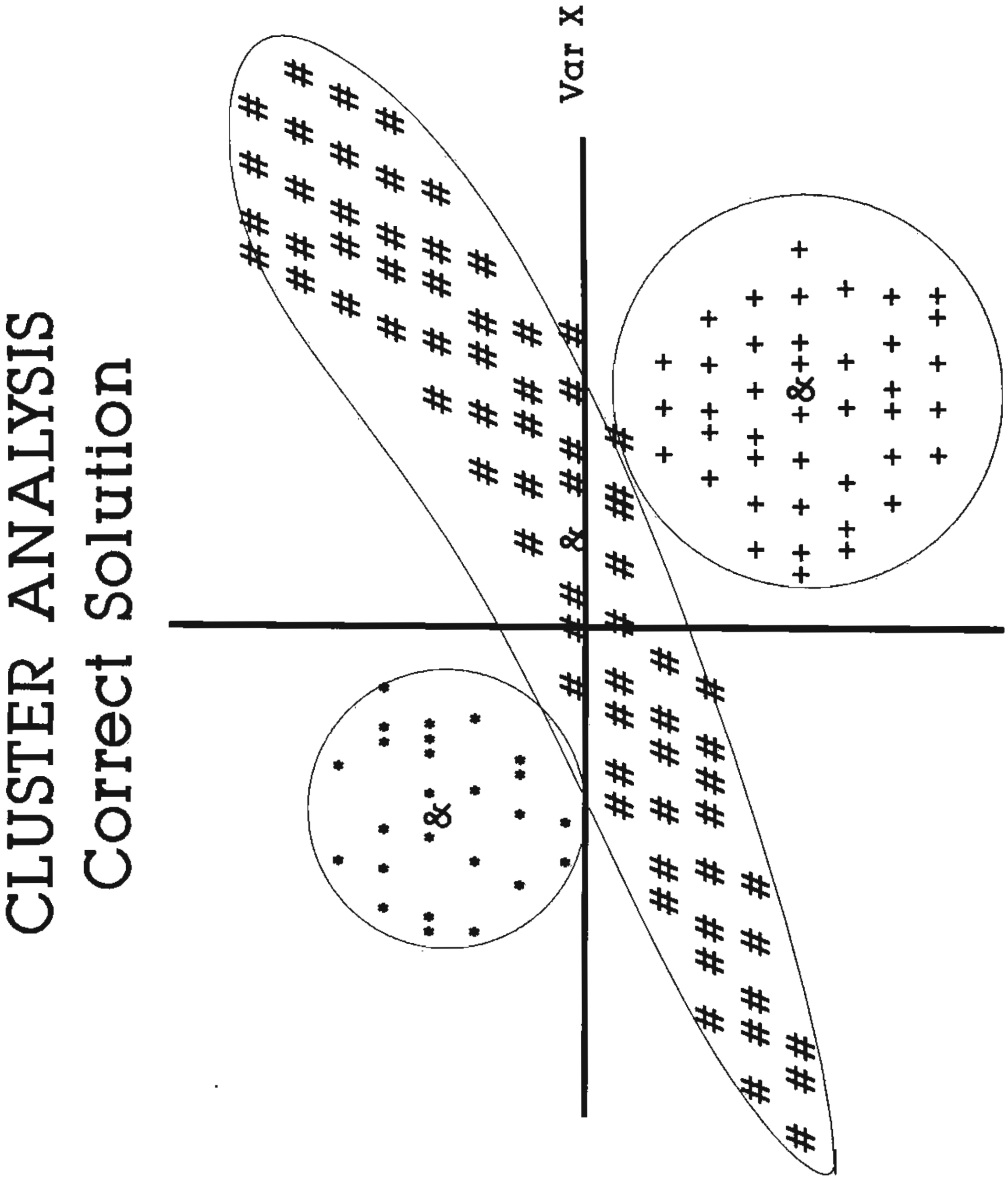


FIGURE 6

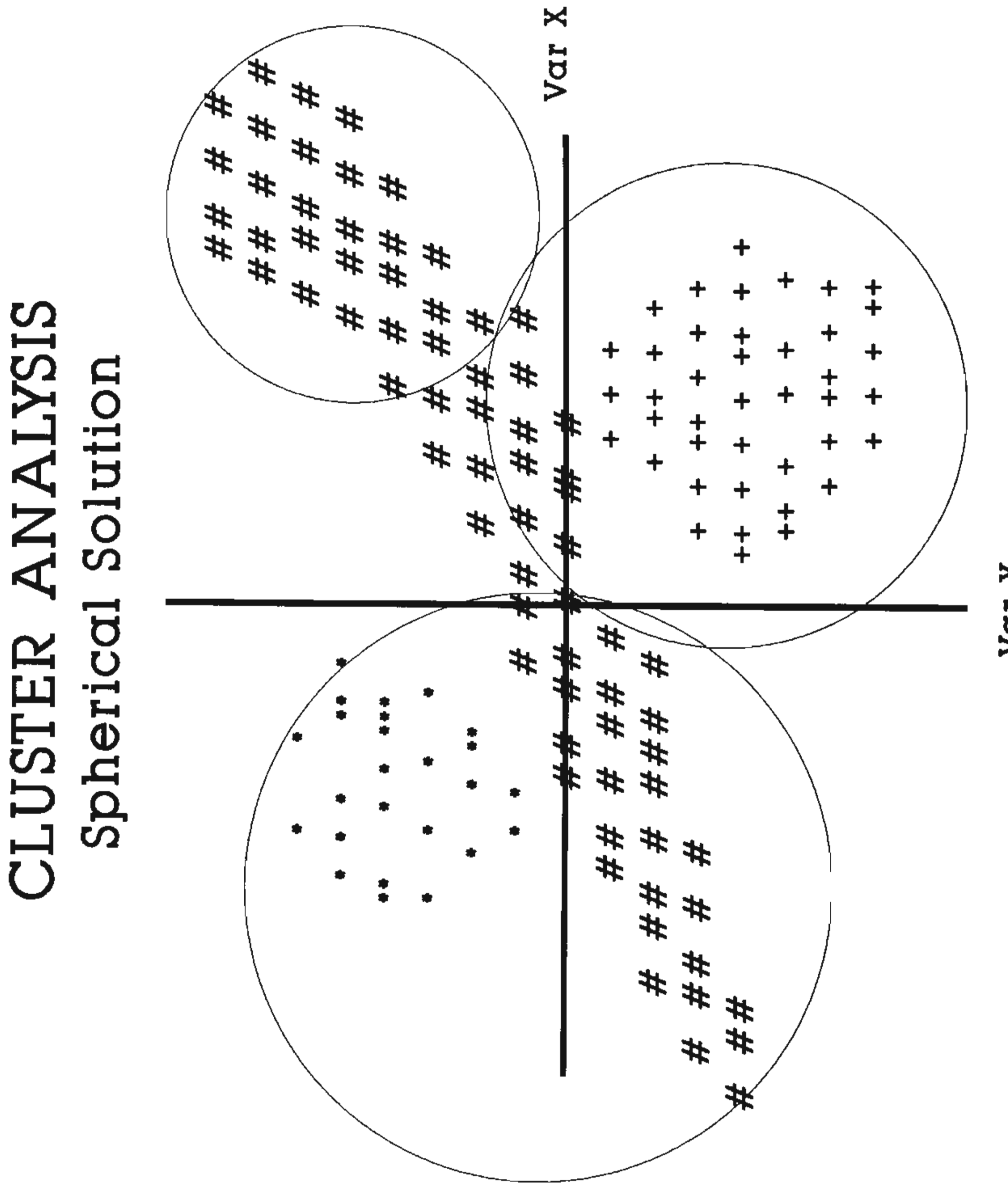


FIGURE 7

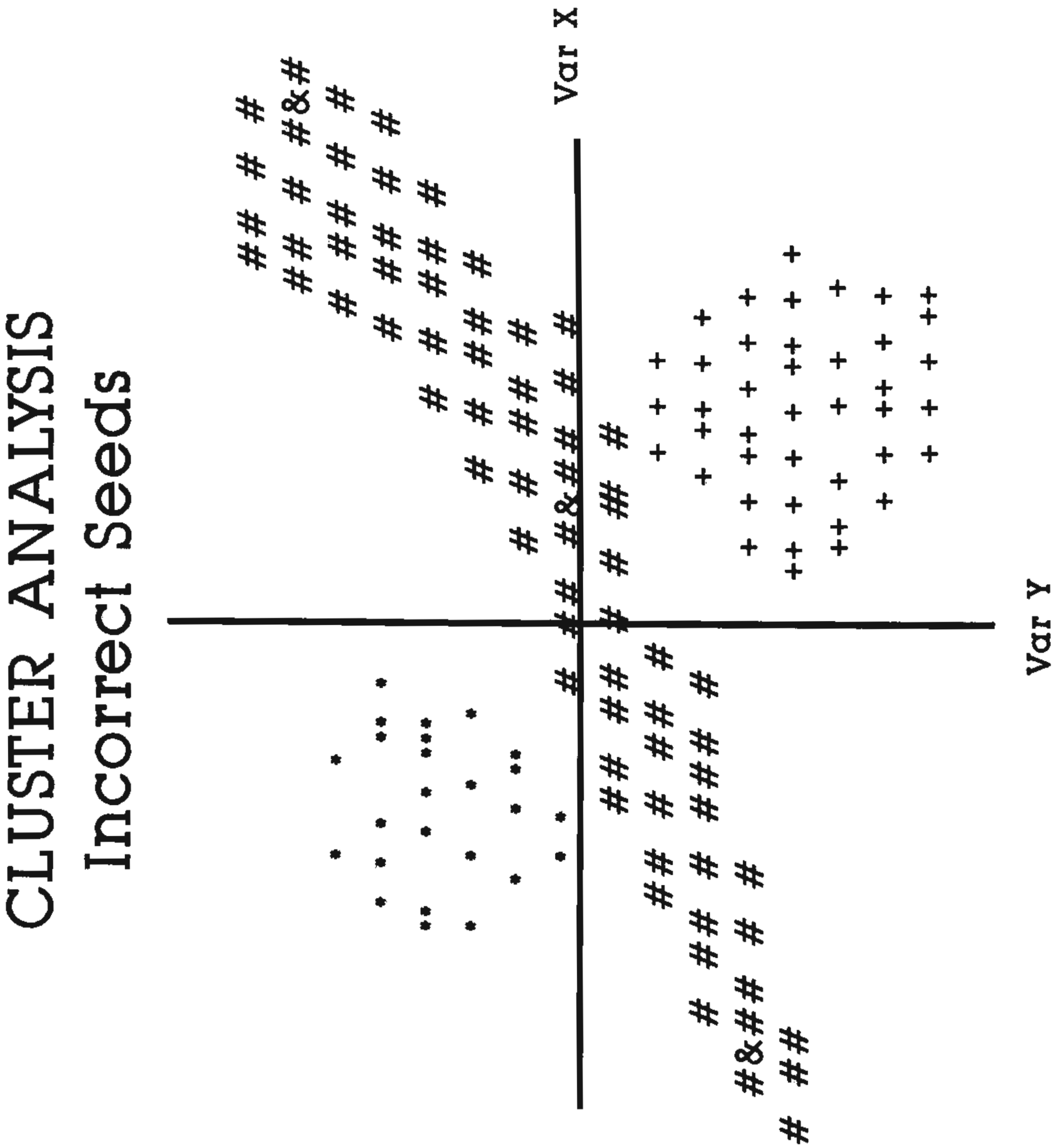


FIGURE 8

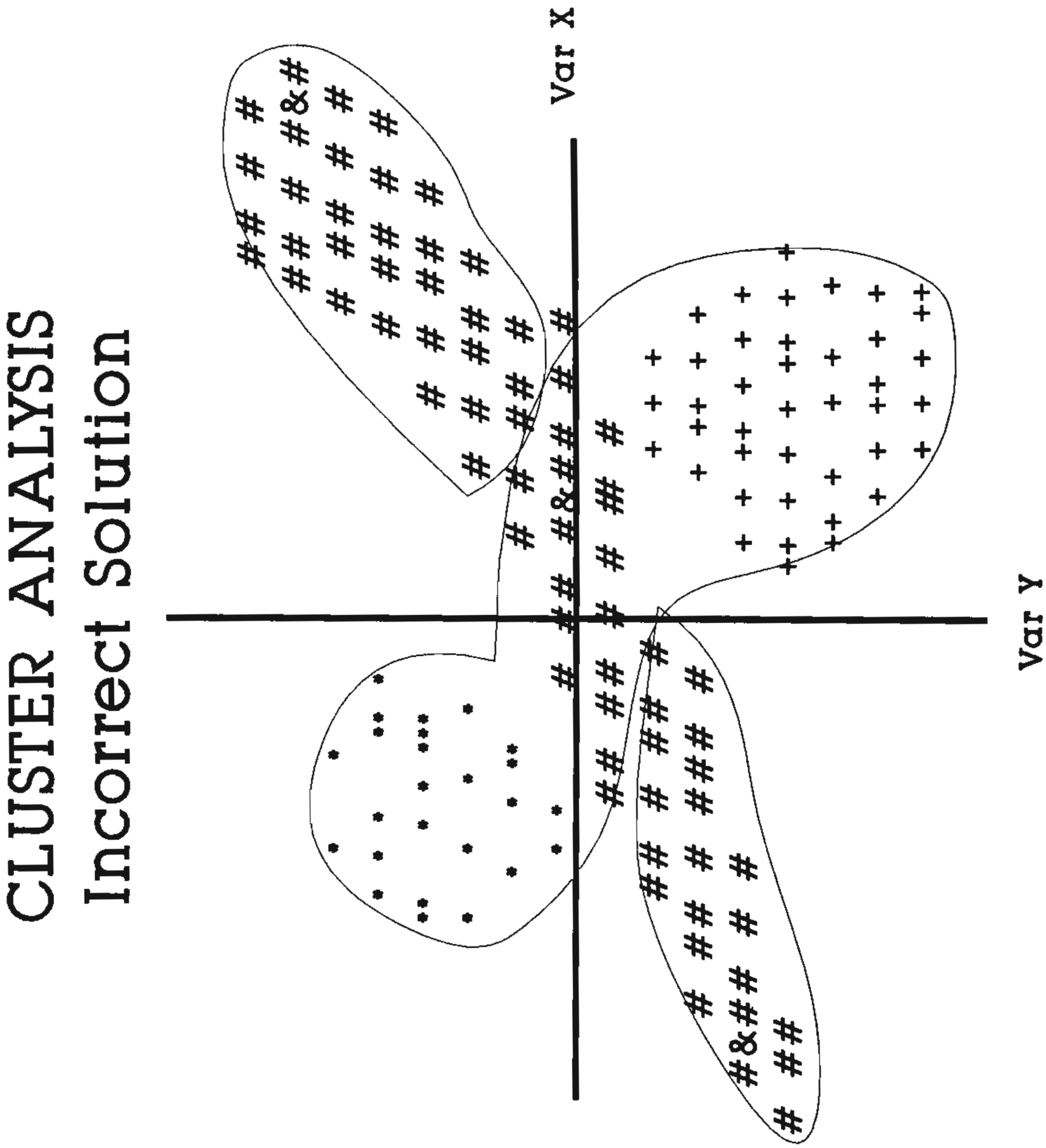


FIGURE 9

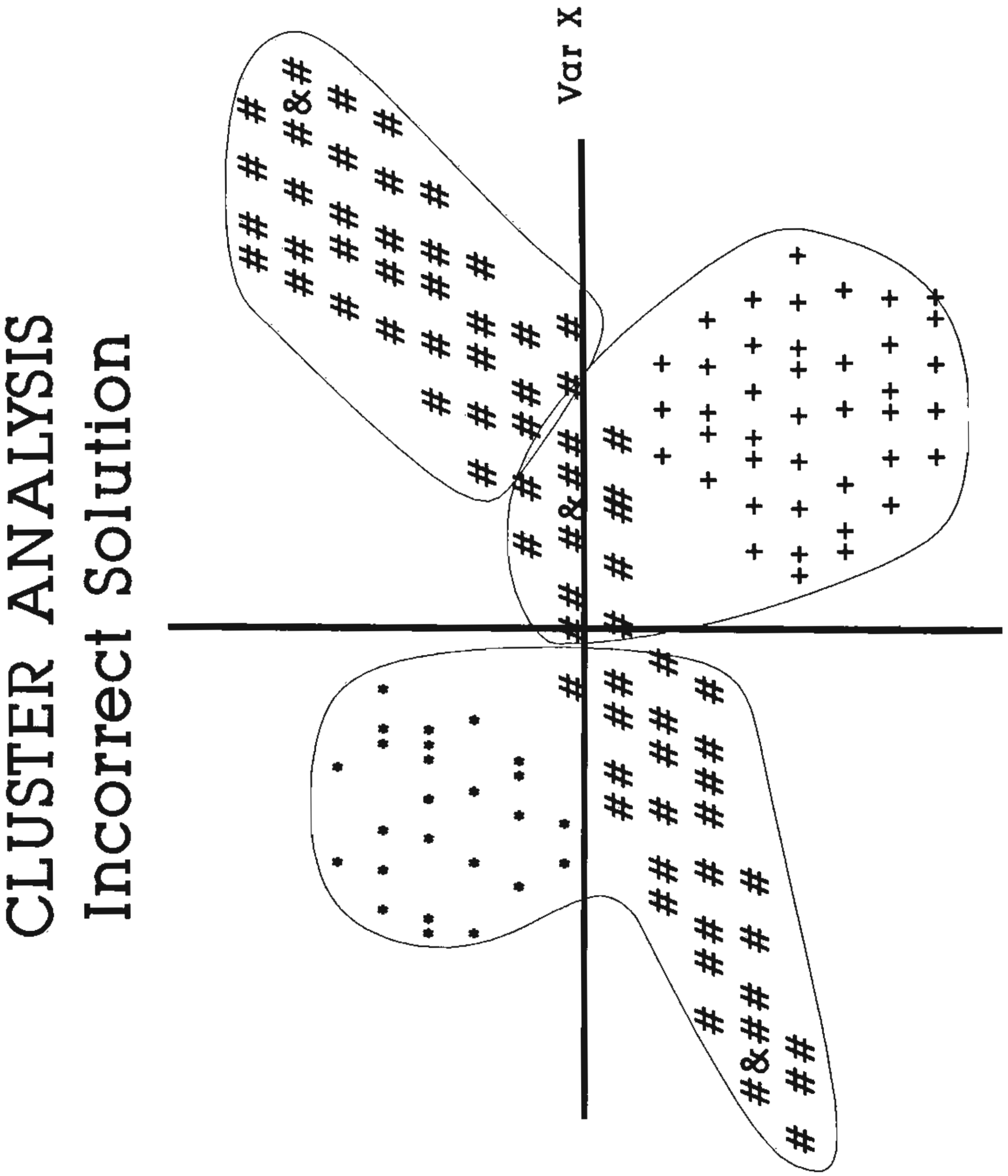


FIGURE 10

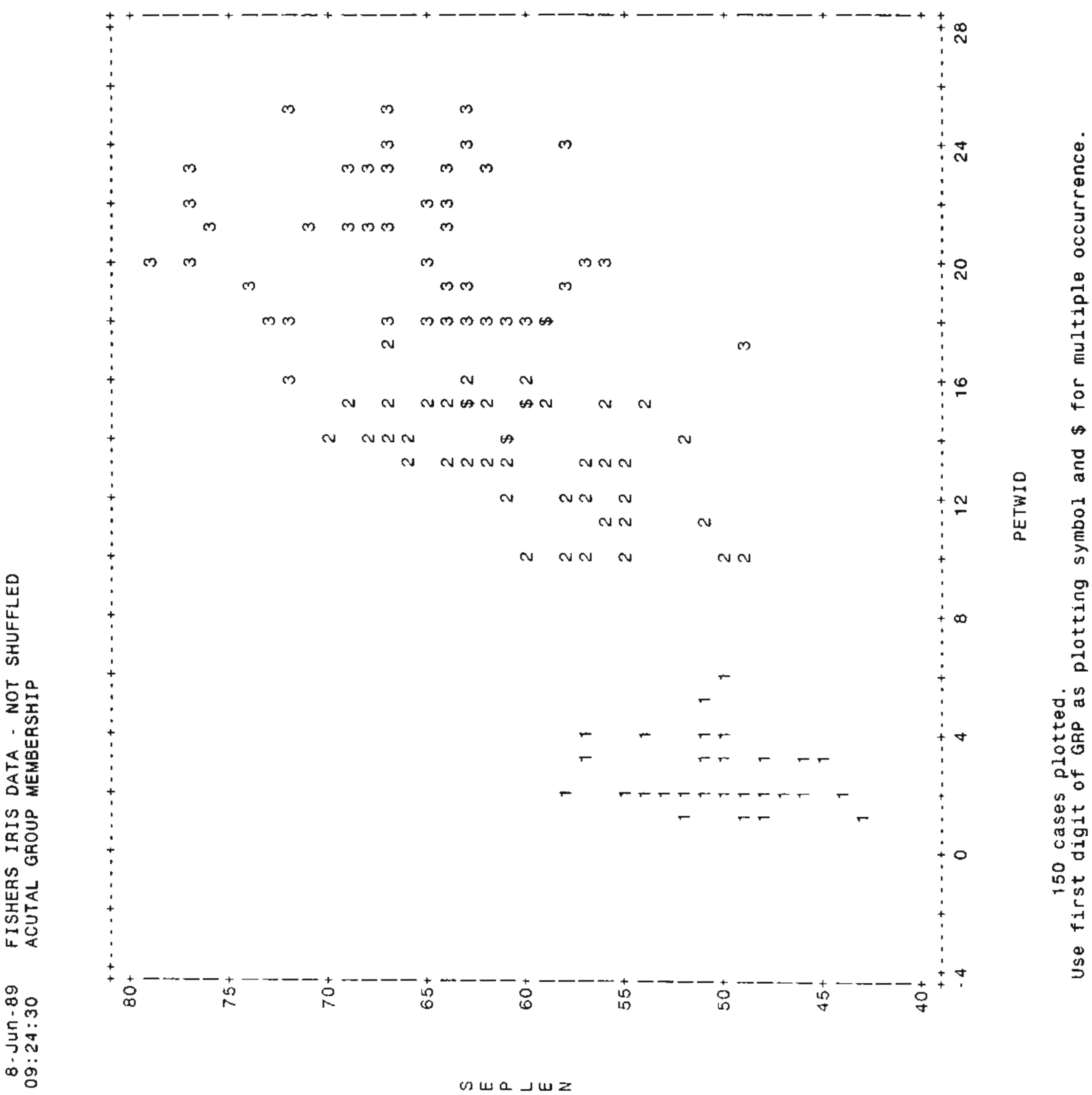
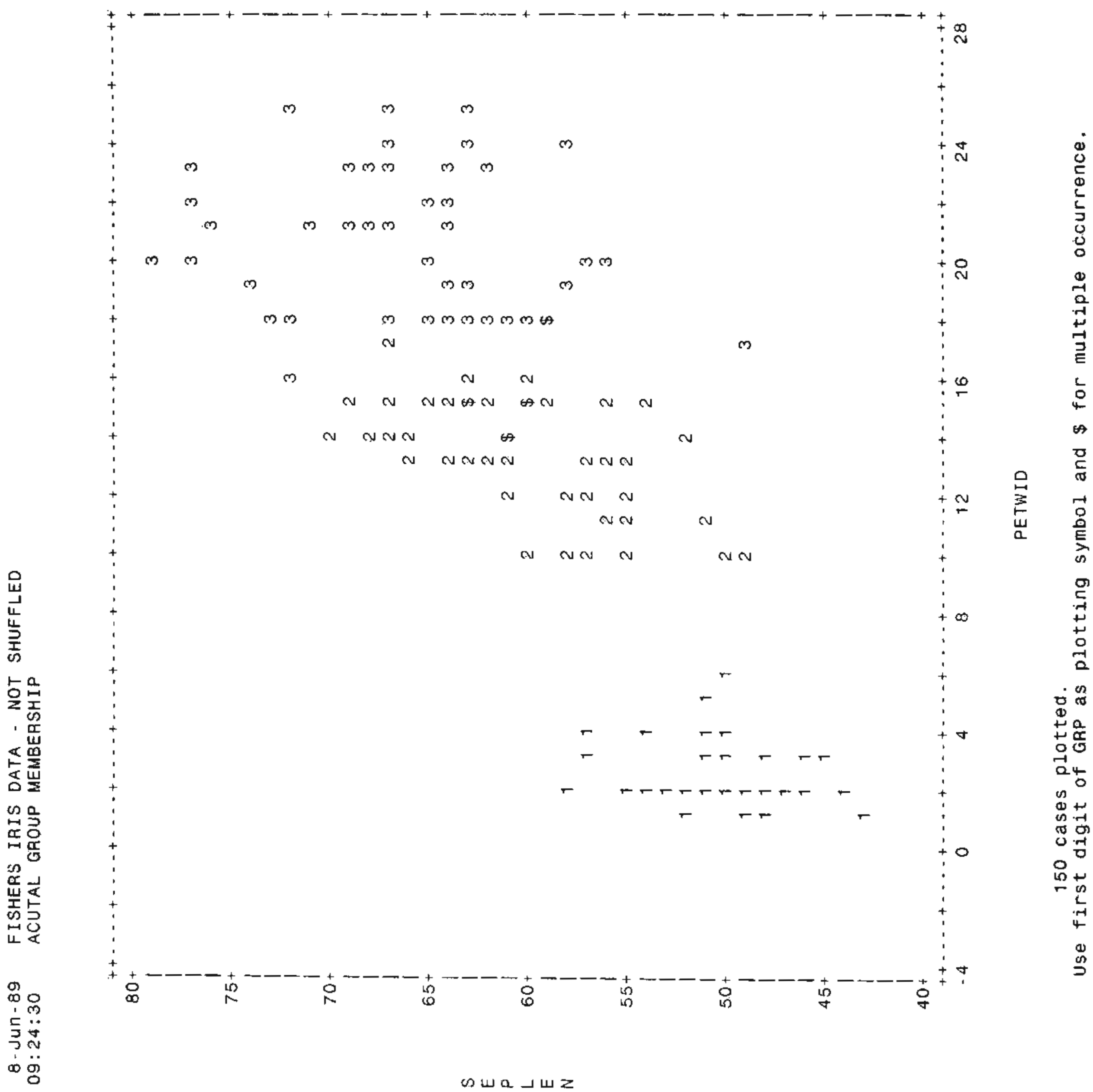

8-Jun-89 FISHERS IRIS DATA - NOT SHUFFLED
09:24:30 ACUTAL GROUP MEMBERSHIP
SEPLEN
80
75
70
65
60
55
50
45
40
-4
0
4
8
12
16
20
24
28
PETWID
150 cases plotted.
Use first digit of GRP as plotting symbol and $ for multiple occurrence.

FIGURE 11


8-Jun-89 FISHERS IRIS DATA - NOT SHUFFLED
09:24:30 ACUTAL GROUP MEMBERSHIP
SEPLEN
PETWID
150 cases plotted.
Use first digit of GRP as plotting symbol and $ for multiple occurrence.

RELIABILITY, DISCRIMINATION, AND COMMON SENSE IN CLUSTER ANALYSIS

Natalie M. Guerlain
POPULUS, Inc.

INTRODUCTION

Perhaps the most common question that arises among users of cluster analysis is: How many clusters actually exist? To arrive at an answer, we use cluster analysis to find several different solutions, each with a different number of clusters. We then compare these solutions to decide which is the "right" one. Applying common sense judgment is the traditional way of making this assessment. While such judgment is critically important, measures of cluster reliability and discrimination provide evidence of cluster resilience and structure. This paper discusses a practical application of cluster analysis using Sawtooth Software's Convergent Cluster Analysis (CCA) software program. The purpose is to:

- Examine clustering issues concerning volumetric data, a kind of data common in marketing research but which call for unique treatment from a clustering standpoint;

- Review some of the procedural decisions in performing cluster analysis; and

- Interpret the resulting cluster solutions using common sense judgment as well as measures of reliability and discrimination.

Common sense judgment assesses the face validity of each solution, based upon existing knowledge of the product category. Employing common sense judgment as a first step, prior to any quantitative assessments of the cluster solutions, enables us to form an independent hypothesis as to which solution best describes the market.

Reliability, defined as consistent cluster recovery, can be measured using several methods: split-sample correlations, reproducibility, and stability. By "reproducibility," we mean the percent of the sample re-clustered together in a subsequent solution that has the same number of clusters. Stability takes reproducibility one step further. By "stability," we mean the degree to which cluster members are re-clustered in solutions with different numbers (k-1 or k+1) of clusters. A diagram of stability provides a hierarchical look at the results of a non-hierarchical clustering method, and can yield dynamic insight into market segmentation.

By "discrimination," we mean the extent to which clusters are different from one another. A measure of discrimination can provide evidence of actual cluster structure: between-cluster differentiation relative to within-cluster similarity. To address the effects of chance, the reproducibility and discrimination of the clusters can be compared to those of clustered random data.

Common sense should come first. It is all too easy to read sense into a wrong solution, just because it appears to be the most reproducible or discriminating. Once an initial common sense assessment is made, however, it may be swayed by compelling evidence of reliability and discrimination in another solution that also makes sense.

THE DATA

In 1988, POPULUS, Inc. conducted a major brand image research study on carbonated soft drinks. Fifteen hundred individuals completed a self-administered Ci2 questionnaire (Ci2 System for Computer-Interactive Interviewing) in 12 central location markets nationwide. Imagery ratings were collected on 40 brands and psychographic, brand consumption, and demographic data were collected.

Cluster analysis was employed to classify respondents into groups based upon patterns of consumption and preference for the 40 brands studied. Respondents were asked to indicate usage of each brand on a 6-point scale:

0 Not in the past year
1 Past 12 months, but not in the past 6
2 Past 6 months, but not in the past month
3 Past 30 days, but drink it rarely
4 Past 30 days, and drink it regularly
5 Past 30 days, and it is a favorite.

This question, typical of marketing research, is one to which respondents can easily relate. It probably violates a fundamental assumption of cluster analysis in that the data are almost certainly not interval scaled. However, a variety of analyses have shown the scale to be remarkably robust. The 40 brands comprise soft drinks of all types and distinct variations exist in brand consumption patterns.

While data for cluster segmentation should normally be standardized, this data set is an exception because it is volumetric. Standardizing is useful with variables that are measured on different scales. Standardizing equalizes variances so that cluster solutions are not dominated by variables with large variances. In this case, however, all variables are measured on an identical scale and brands with greater variances are the more important ones. These differences in variances, where they occur, act as "natural" weights in the clustering process.

CLUSTERING PROCEDURE

CCA utilizes the convergent k-means method of clustering. This clustering method uses one or more methods to select "starting points" (respondents), which act as initial (prototypical) "cluster centers." Clusters are formed by assigning each respondent to the cluster with the most similar starting point. Then the starting points are replaced by improved points (cluster averages) to which respondents are reclassified. This iterative process continues until no respondents are reclassified. In clustering the soft drink data, the maximum number of iterations (100) was allowed, following suggested procedure.

The quality of a solution is highly dependent upon the adequacy of its starting points. To account for this, CCA is designed to replicate each solution several times using different starting points, and to utilize various methods to select these starting points. In this research, five replications were requested in all clustering runs, using a combination starting point strategy. Clustering was performed on the full sample (N = 1502), for which 3-cluster to 8-cluster solutions were produced.

ANALYSIS OF SOLUTIONS

COMMON SENSE

The soft drink solutions were first examined for common sense, an assessment otherwise known as "face validity" or "expert judgment." Each cluster was profiled according to two types of variables. Basis variables, the 40 brands, showed each cluster's profile of brand consumption. Some clusters clearly preferred specific types of soft drinks, while others shared preferences of moderate degrees. Descriptive variables, those not included in the segmentation process, were valuable for filling out each cluster's profile in terms of age, sex, preferred type of soft drink (diet or regular), and overall volume of soft drink consumption.

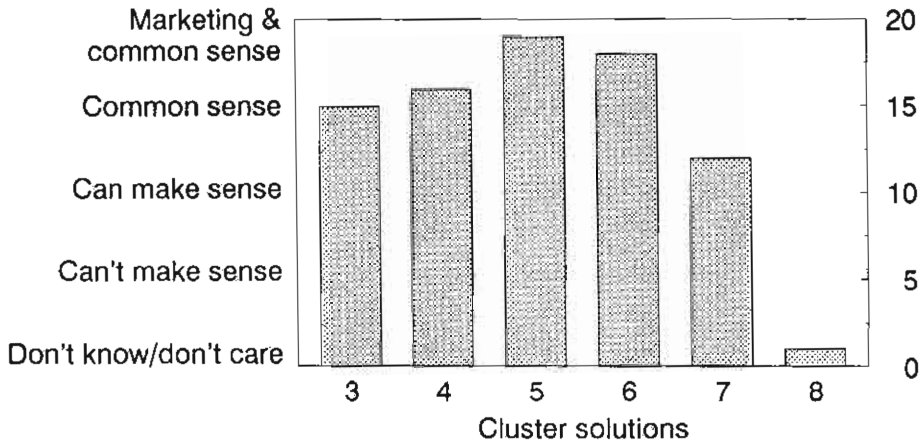
In examining the soft drink clusters, the "right" solution proved somewhat elusive from a purely common sense standpoint because many of the solutions made sense. The external variables, in particular, not only confirmed the basis variable profiles, but strongly aligned with existing knowledge of soft drink consumer classifications. Nevertheless, the solutions could be assessed. A Common Sense Scale was developed to rate them:

Common Sense Scale	
It makes a great deal of sense: marketing sense as well as common sense	20
It makes common sense	15
I can make sense of it	10
I can't make sense of it	5
Don't know / doesn't matter	0

The combined assessment of five research professionals showed the 5-cluster solution to make the most sense (Figure 1). The 3-cluster solution was too condensed; it made sense but was too simplistic. The 5- and 6-cluster solutions were much more informative, but the 6-cluster solution was less intuitive. The 7-cluster solution was the point at which too much splintering arose, obviating consideration of the 8-cluster solution as well.

Figure 1

COMMON SENSE ASSESSMENT



RELIABILITY

Reliability and discrimination measures supplement traditional common sense judgment by providing evidence of cluster resilience and structure. As we further investigated the soft drink solutions, evidence accumulated favoring selection of a different solution.

Split-sample Correlations

One measure of cluster reliability is split-sample correlation. If clusters of separate subsamples are highly similar, then they may presumably reflect inherent market structure rather than grouping just due to chance. Using the basis variable profiles of each cluster, we measured the degree of similarity of each of the k clusters in one subsample to each of the k clusters in the other subsample.

Split-sample correlation is a four-step process: randomizing the cases, splitting the sample, clustering each half, and then correlating the halves. The data for the correlation analysis are clusters' mean deviations from the cluster solution's grand mean for each brand, as illustrated in Table 1. The correlations are across the 40 brands. The profile of each cluster in subsample A is correlated with the profile of each cluster in subsample B.

Table 1
Illustration of Data for Correlation:
Cluster Mean Deviations from Grand Mean
(3-cluster Solutions)

	Subsample A Clusters			Subsample B Clusters		
	1	2	3	1	2	3
Brand 1	-1.2	0.3	2.2	0.3	3.6	1.7
Brand 2	-0.1	0.9	-1.0	-0.1	-1.1	0.9
Brand 3	-0.5	1.2	-0.7	-0.8	-0.4	1.2
:	:	:	:	:	:	:
:	:	:	:	:	:	:
Brand 40	-0.3	0.8	-0.6	0.7	-0.6	-0.2

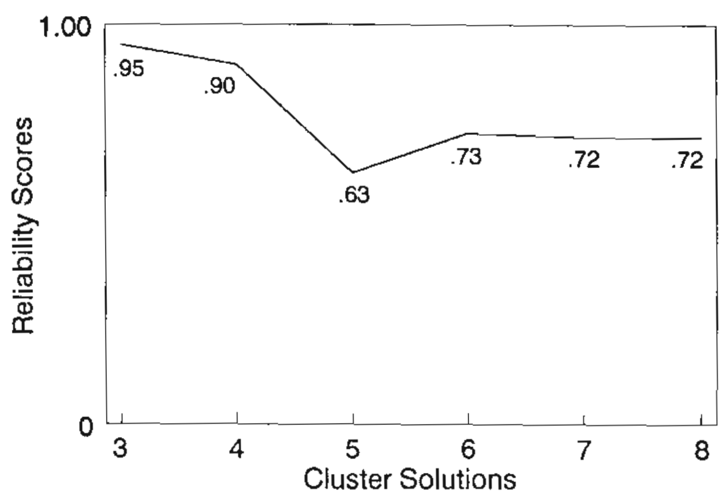
Correlating the 3-cluster subsamples resulted in a 3x3 correlation matrix (Table 2). High overall correlation was evident from the fact that each of the A clusters was highly correlated with one and only one of each of the B clusters. Averaging the selected coefficients resulted in a 3-cluster reliability score of .95.

Table 2
Split-sample Correlation Matrix
(3-cluster solutions)

	B1	B2	B3
A1	**.9414**	-.5204	-.1433
A2	-.1319	-.6823	**.9645**
A3	-.6048	**.9747**	-.7093

In solutions with more clusters, lower reliability would be expected. But split-sample correlations for the soft drink solutions indicated that the 5-cluster solution had the lowest overall reliability (Figure 2). This was the first evidence to call into question the common sense judgment regarding the 5-cluster solution as the "right" one.

Figure 2
SPLIT-SAMPLE CORRELATION RELIABILITY



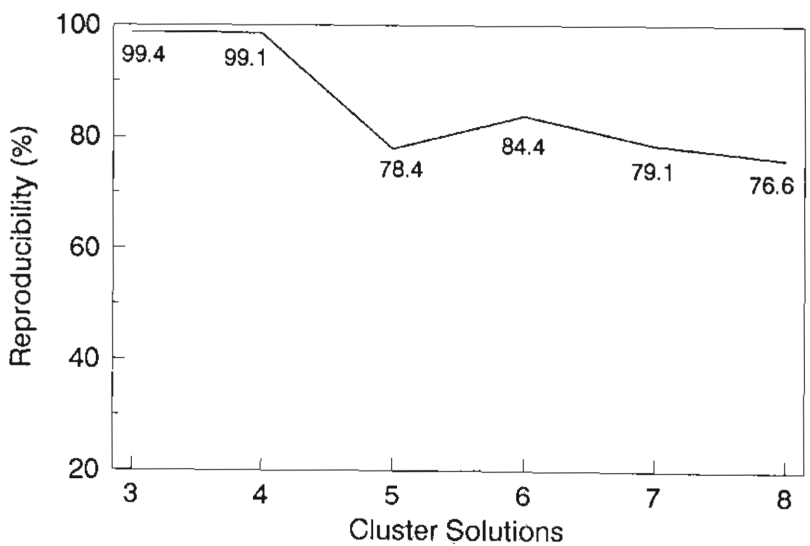
Reproducibility

While correlation compares the variable profiles of clusters from different samples, reproducibility compares clusters produced from clustering the full sample repeatedly. CCA is designed to perform several replications of each solution. It then counts the number of respondents who land in the "same" clusters in each pair of replications (expressing the result as a percentage of the total sample) and selects the replication that is most reproducible.

Normally, reproducibility declines as the number of clusters increases, due to increased degrees of freedom. The reproducibility of the various soft drink solutions is shown in Figure 3, where two deviations from this pattern are apparent. The 4-cluster solution is no less reproducible than the 3-cluster solution. More striking is the 6-cluster solution: its reproducibility "elbows" up from the 5-cluster solution. This elbow in particular indicates that something may be "happening" in terms of real classification structure. It also lends further evidence against the 5-cluster solution favored by common sense judgment.

Figure 3

SOFT DRINK REPRODUCIBILITY



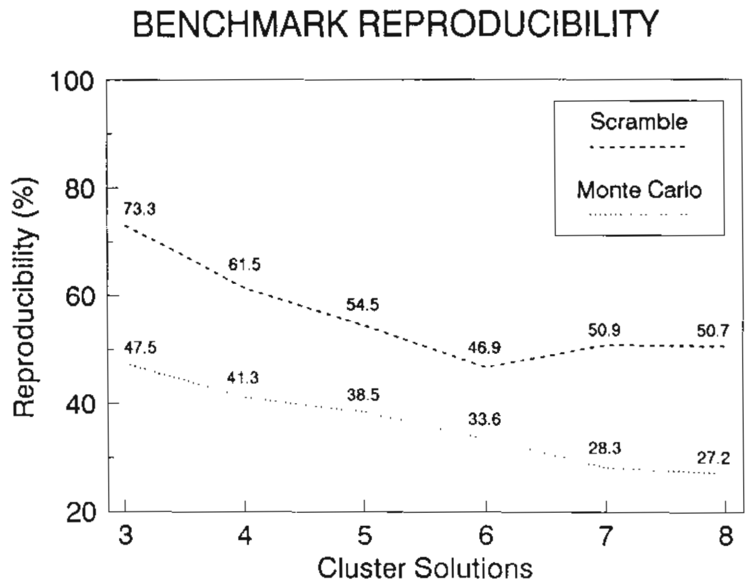
Benchmark Reproducibility

When clustering purely random data, not only will the algorithm discover some sort of pattern and produce clusters, but these clusters will exhibit some degree of reproducibility. Therefore, reproducibility levels should be examined relative to the amount of reproducibility due to chance. Two data sets were developed for this purpose.

<u>Monte Carlo</u>. A data set composed of uniformly distributed random numbers similar to the soft drink data set (40 variables, scaled 0 to 5) was generated and clustered. The results are shown by a dotted line in Figure 4.

Scramble. A more direct means of comparison was to add error to the soft drink data: scrambling the within-respondent patterns that provide the basis for structure. The values for each variable were scrambled (randomly permuted) across the 1502 cases. This process resulted in a data set in which the means and deviations of each of the variables remained the same, but in which any real structure was destroyed. The Scramble reproducibility results are shown by a dashed line in Figure 4.

Figure 4



The reproducibility of each of the soft drink solutions indexed to its scrambled counterpart revealed that the greatest differences occur in the 4- and 6-cluster solutions (Table 3). The 6-cluster solution in particular is 79% more reproducible, confirming the integrity of the 6-cluster "elbow."

Table 3
Solution Comparison:
Reproducibility

# Clusters	Soft Drinks	Scramble	Index
3	99.4%	73.3%	135
4	99.1	61.5	161
5	78.4	54.5	143
6	84.4	46.9	179
7	79.1	50.9	155
8	76.6	50.7	151

The levels of reproducibility in these "garbage" data sets are striking, demonstrating the power of cluster analysis to produce similar clusters repeatedly, in spite of our best efforts to remove any real structure from the data. It also demonstrates the need for benchmarks when analyzing cluster solutions.

Scramble is more reproducible than the purely random data, presumably because the variances of its "variables" are less equal. Incidentally, the reproducibilities of standardized versions of these two benchmark data sets are much more similar, as might be expected (not shown).

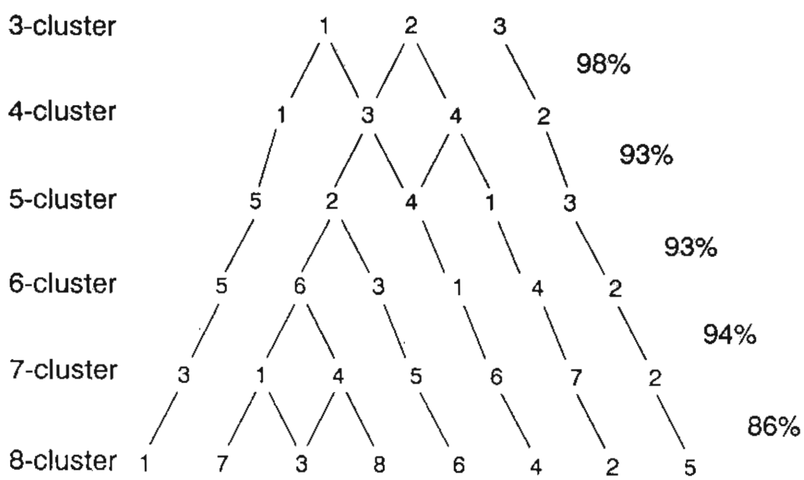
Stability

Convergent k-means is a "non-hierarchical" clustering method, meaning that its k+1-group solution is developed independently of its k-group solution. Still, a somewhat hierarchical relationship may emerge from a non-hierarchical method. Stability analysis enables us to learn 1) which individual clusters are stable and which are not, and 2) in which solution most clusters appear to stabilize across the board.

The soft drink clusters are largely hierarchical, as the diagram of the 3-cluster through 8-cluster solutions (Figure 5) shows. Most of the 3-cluster solution's cluster #3 membership remains together (it regroups repeatedly down the far right of the diagram). The 3-cluster solution's cluster #1 is also highly independent of the rest of the sample.

Figure 5

STABILITY



Reading common sense descriptions into this picture clarifies at which level (in which cluster solution) most of these highly hierarchical clusters stabilize. Sex, age, and other descriptive variables reveal sharpened profiles at either the 5- or 6-cluster levels relative to the 3- or 4-cluster levels.

It may be recalled that both the 5- and 6-cluster solutions make a great deal of sense. It is evident from the stability diagram that they are quite similar, except that the members of the 5-cluster solution's largest cluster (#2) are separated into two clusters in the formation of six clusters. Despite the apparent similarity, we have learned that the 6-cluster solution is more reliable.

The percentage of the sample which is re-clustered together from one level to the next is also shown. The sample is highly stable overall.

DISCRIMINATION

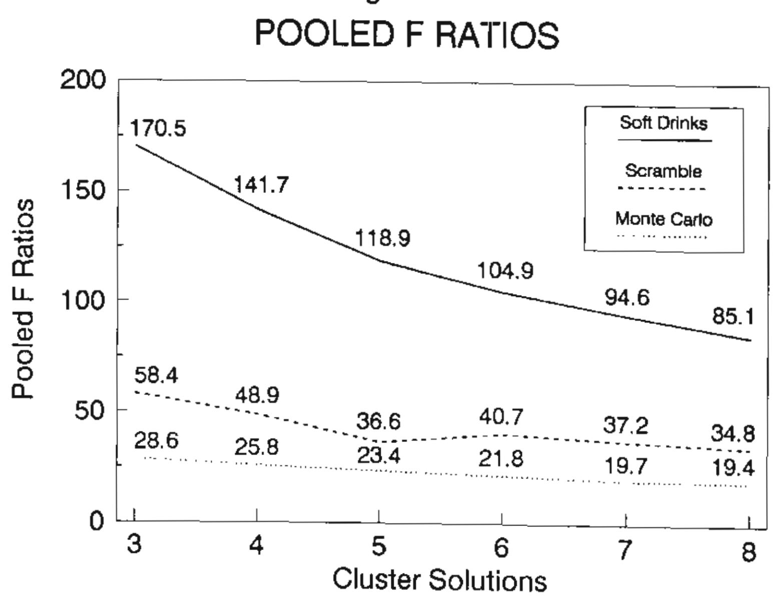
Clusters must be discriminating as well as reliable. That is to say, the members of one cluster should be similar to one another and different from those in other clusters.

CCA's Measure of Discrimination: The Pooled F Ratio

In each cluster solution, CCA calculates an F ratio for each variable. These F ratios, the "mean square between clusters" divided by the "mean square within clusters," indicate the relative amount of difference among clusters due to those variables. Collectively, these differences are summarized in a "pooled F ratio," an overall measure of discrimination among clusters.

Like reproducibility figures, pooled F ratios gain meaning when compared to benchmark data. The comparative results in this application provided evidence of cluster structure in the soft drink data. Because levels of discrimination were so high in every case, however, indexing these results did not provide insight into the superiority of any one solution.

Figure 6



Discrimination: External Data Validation

Statistical tests cannot be performed upon the F ratios produced by cluster analysis. However, they can be performed upon variables not used in the clustering. This helps to validate the differences between clusters. In this study, t-tests and analyses of variance revealed significant differences in almost every instance.

CONCLUSIONS

This analysis of the soft drink brand consumption and preference clusters revealed cluster groupings that made sense both intuitively and when analyzed on outside variables, and were so discriminating as to make it difficult to decide which solution was "best."

A common sense analysis focused initial attention on the 5-cluster solution, with the 6-cluster solution a close second. Split-sample correlation and reproducibility analyses provided strong backing for the 6-cluster solution instead.

Benchmark comparisons employing random data clustered under the same conditions as the research data were found to be useful in assessing reliability and discrimination.

Stability analysis revealed very high overall stability, as well as insight into which individual clusters were more independent, and which cluster solutions were more stable.

Volumetric data is one of the rare instances in which data should not be standardized in preparation for cluster analysis. Using unstandard data holds the potential for allowing more discriminating variables to have greater influence in a "natural" weighting process, though at the risk of producing solutions overly dominated by certain variables. The existence of significant and sensible differences among the resulting clusters on external criteria supported the viability of clustering these non-standardized data.